THE VOICED-VOICELESS DISTINCTION
AND
ASSIMILATION OF VOICE IN DUTCH

I.H. SLIS

THE VOICED-VOICELESS DISTINCTION AND ASSIMILATION OF VOICE IN DUTCH

CIP-GEGEVENS KONINKLIJKE BIBLIOTHEEK, DEN HAAG

Slis, Iman Hans

The voiced-voiceless distinction and assimilation of voice in Dutch / Iman Hans Slis. - [S.l. : s.n.] (Helmond : WIBRO)
Proefschrift Nijmegen. - Met lit. opg. - Met samenvatting in het Nederlands.
ISBN 90-9000845-4
SISO * 837.2 UDC 803.931-4
Trefw.: Nederlandse taalkunde ; fonetiek.

THE VOICED-VOICELESS DISTINCTION

AND

ASSIMILATION OF VOICE IN DUTCH

PROEFSCHRIFT

TER VERKRIJGING VAN DE GRAAD VAN DOCTOR IN DE LETTEREN AAN DE KATHOLIEKE UNIVERSITEIT TE NIJMEGEN, OP GEZAG VAN DE RECTOR MAGNIFICUS PROF. DR. J.H G.I. GIESBERS, VOLGENS BESLUIT VAN HET COLLEGE VAN DEKANEN IN HET OPENBAAR TE VERDEDIGEN OP DONDERDAG 21 FEBRUARI 1985 DES NAMIDDAGS TE 4 00 UUR.

DOOR

IMAN HANS SLIS

GEBOREN TE BANDOENG

Druk Dissertatie Drukkerij Wibro, Helmond

ISBN 90-9000845-4

Promotores:
Prof. Dr. W.H. Vieregge
Prof. Dr. A. Cohen

CONTENTS

PREFACE

The work reported upon in this dissertation covers a long period of research (from 1962 till the present day) conducted in two different institutes. Consequently many people have influenced the course of this research, for which I am grateful.
I took the first, hesitant, steps in the field of phonetics at the IPO (Instituut voor Perceptie Onderzoek, Eindhoven) under the guidance of Toon Cohen. To him I owe my basic knowledge of phonetics and my interest in fundamental research. His teaching consisted not of dealing out instructions but of asking why.
Although I am indebted to nearly all the members of the IPO for their discussions, recommendations and corrections in their capacities of members of a publication committee or a group giving instrumental advice, or simply as interested colleagues, nevertheless I should like to mention some of these people by name.
Hans ´t Hart, who was my room-mate in the first years at the IPO and who introduced me to the first synthesizer, the IPOVOX-I; Ab van Katwijk with whom I worked among other things on speech synthesis rules before his premature death; Sieb Nooteboom, with whom I collaborated in numerous timing experiments; Jan Eggermont, who functioned as a support in difficult situations; Herman Muller, who helped me in implementing my first synthesis by rule program; Lei Willems, whose knowledge of instrumentation was unrivalled.
In 1976 I left the IPO for the IFN (Instituut voor Fonetiek, Nijmegen). I also owe my IFN colleagues an enormous debt of gratitude for their help, interest and support in my work. University regulations forbid me to mention names of members of the faculty in the acknowledgements; suffice it to say that what I have said above of the IPO-staff holds for my IFN colleagues too.
Moreover I thank all the investigators who have inspired me in my work by their stimulating books, articles, papers and discussions. Only some of these people are mentioned in the reference lists of my articles and this dissertation. There are three names that deserve a special mention here: Professor Eli Fischer-Jørgensen, who has supplied the phonetic world with an enormous amount of detailed information on many different languages, and Professors Arthur Abramson and Leigh Lisker, who I regard as my antipodes in the field of the voiced-voiceless distinction; without their experimental evidence, my understanding of the topic would have been far from complete.
Last but not least, I want to mention my wife Jenny and my daughters Anneke and Margreet. Doing research and especially getting the right ideas at the right times, is highly dependent on the mental state of the investigator. For this, a suitable domestic background is a most important condition.

CHAPTER 1

PROLOGUE

1.1 INTRODUCTION

Differences between the classes of speech sounds consisting of /p,t,k,s,f,x/ and of /b,d,g,v,z,ɣ/, which will be called voiceless and voiced obstruents respectively, are an old issue that plays an important role in phonetic research. With respect to the question what the essential features are that constitute this difference various possibilities have been proposed in the literature from the 17th century up to now.

In the opinion of Petrus Montanus (1635), an early Dutch linguist, the presence or absence of activity of the vocal cords is essential in the opposition between voiced and voiceless obstruents; in his description he divides the plosives ("smoorvormen" = smothered forms) into those with a sound source ("stemgat" = voice opening) at the larynx ("clinkgat" = sonorous opening in the throat) and those with a noise source ("ruisgat" = noise opening) which can be positioned at the constriction ("scheideur" = dividing-door) in the vocal tract. The constriction is closed during plosives, but the noise source becomes evident in the adjoining segments ("cleefsels" = cling-ers) as a sucking or puffing sound. The fricatives are divided into those with one sound source, viz. a noise source and those with two sound sources, viz. a noise source and a sonorous source . Montanus also observed that the voiced sounds require less effort than the voiceless ones. In voiced sounds the air can flow smoothly and fast through the constriction, while in voiceless sounds the air meets an obstruction so that the applied force is broken.
Petrus Montanus also described regressive assimilation of voice in stop-stop clusters; he observed that the English speech sound /g/, which does not exist in the sound system of Dutch ("... wiens plaets in mijn Letterorde ledich stont..."), was perceptable instead of /k/ in words like "hekboot" and "slikboot", for the same reason as /t/ changes in /d/ in "zitbank" and /p/ changes in /b/ in "loopbaan".
About a century later, Lambert ten Kate (1723) described the plosives D, T, B, P and K as bluntly broken off, without any after sound, the five of which are called mute ("... bot-af-brekende, zonder eenig nageluid ... welke vijf men stomme noemt ..."). According to Ten Kate voiced and voiceless obstruents differ from each other with respect to their formation and pronunciation in no other aspect than in the sharpness of cut-off (which term would appear to refer to the oral closing of the preceding vowel); of these obstruents V, B, D, Z, G are the soft and F, P, T, S, CH the sharp ones ("...ten opzigte van hun vorming en uitspraak van den ander niet verder verschillen dan in de scherpte van afsnijding van welke de ...(V, B, D, Z, G)... de zagte en de ... (F, P, T, S, CH)... de scherpe zijn."). Obviously he did not notice the presence or absence of voice in the voiced or voiceless obstruents.
Ten Kate also paid attention to assimilation of voice; he speaks of the inclination to an agreeable pronunciation (euphonia) ("... de trek tot

een gevallige uitspraak (euphonie) ..."). A sharp consonant will change an adjoining soft consonant into a sharp one. He did not observe cases in which a sharp consonant became soft. It must be pointed out however that Ten Kate defined final obstruents to be "soft by nature"; nowadays, final obstruents are regarded to be voiceless (and ought therefore to be called sharp) because of a final devoicing rule.
A more recent well-known Dutch precursor in phonetics is F.C.Donders. More than a century ago he already used oscillograms for phonetic research. Donders (1865) related different terminologies to each other; he spoke of obstruents which are "sonorant" ("luidend": produced with a narrow glottis, and consequently voiced) which can also be defined as "hard" ("hard"), and obstruents which are "whispery" ("fluisterend": produced with a wide glottis, and consequently voiceless) which can also be defined as "weak" ("week").
In the same period Cohen-Stuart (1872) warned for perceptual errors by stating: how writing may deceive our hearing!... and still nearly nobody notices that we write f and s over and over in places where we pronounce v or z, and the reverse; yes, even if our attention is drawn to it, many a man will experience that it is difficult to free himself from the influence writing exercises on his conception ("... hoe toch het schrift ons bedriegen kan!... en toch bemerkt bijna niemand, dat wij telkens f of s schrijven, waar wij v of z uitspreken, en omgekeerd; ja ook na er opmerkzaam op gemaakt te zijn zal het menigeen moeite kosten zich los te maken van de invloed, dien het schrift op zijne voorstelling uitoefent.").
In the first half of the present century Zwaardemaker and Eykman (1928) placed the difference in articulatory effort in the foreground by pointing out that the articulation can take place in a robust or less robust way ("... kan forscher of minder forsch plaats hebben."). If the difference in "energy" is evident one speaks of tense or tight vs. loose (reference to Sweet 1908) , or fortis vs. lenis (reference to Sievers), in which the voiceless one is tight or fortis and the voiced one is loose or lenis. Zwaardemaker and Eykman also paid attention to assimilation of voice; they used a system in which six degrees of voiceless/voiced can be distinguished.

In the first half of our century, assimilation of voice has been a point of discussion in numerous articles, but the terminology was used in an inconsistent way.
At first, the term assimilation was used for historical sound changes, indicating that pronunciation is subject to an evolutionary process, as well as for context-dependent sound changes, indicating that one speech sound changes under the influence of other speech sounds in a particular context. It is this type of assimilation which is the subject of our study.
Secondly, various terms were used to indicate assimilation e.g. sandhi (Viëtor 1898), or sound-junction (Sweet 1908). With respect to a detailed description regarding direction of assimilation, terminology was often confusing. What we call regressive vs. progressive assimilation was also called "rückschreitend" (backward), or "vorgreifend" (anticipatory) vs. "fortschreitend" (forward), "verweilend", or "beharrend" (persevering) assimilation (Jespersen 1904, 1912), "anticipatory or proleptic" vs. "persevering or analeptic

assimilation" (Michels 1957), and "anticipatory" vs. "carry-over coarticulation" (e.g. Schouten and Pols 1979).

In the descriptions of assimilation of voice in Dutch in this period, a great number of factors affecting assimilation can be found. Meinsma (1958) listed ten different influences. The fact that so many factors seem to play a role, the confusing terminology, as well as the assumed perceptual prejudices of at least some of the investigators probably explains the inconsistencies found in the older descriptions of assimilation.

In the present-day literature the various views on the most characteristic features of the voiced-voiceless distinction are still a point of discussion, in which acoustical, perceptual, kinesthetic and linguistic arguments play a role. It is often not clear which of the approaches mentioned is most relevant. The terms voiced-voiceless and sonorant-whispered, for instance, are presumably from an articulatory/acoustical/perceptual origin, fortis-lenis and tight-loose, on the other hand, from an articulatory/kinesthetic origin. In the early days of modern phonology, when a phoneme was regarded to consist of a bundle of distinctive features, the old, existing terms were adopted. These were assigned a descriptive linguistic function, in which word discrimination played a fundamental role.
The fact that various terms based on different aspects were available, could be (partly) the cause for the introduction of different distinctive features, e.g. voiced-voiceless and fortis-lenis (or tense-lax), in cases where one feature might have been sufficient. This redundancy in features for the voiced/voiceless distinction was used by some investigators to indicate the degree of assimilation.

This volume deals with the question in which way the distinction between voiced and voiceless obstruents can be described phonetically in Dutch and whether there are phonetically based arguments to use one or more distinctive features for an adequate description. These matters will be discussed with respect to single, intervocalic obstruents and obstruents in two-consonant combinations in which assimilation may or may not occur.

1.2 METHODOLOGICAL ASPECTS

The views of classical phonetics referred to above, were mainly based on introspective/intuitive experiences of the speech of the investigators themselves and on the auditory impressions of the investigators of the speech of other people. These data were supplemented with measurements of articulatory events, which were obtained in often ingenious ways.
In more recent years, after the 1940's, phonetic research, and consequently research on the voiced-voiceless distinction, has expanded as a result of developments in electronics. Tools became available for the recording and analysis of the speech signal. As a consequence an acoustical description was added to the introspective and mostly articulatory descriptions. In addition, tools were designed for manipulation of recorded speech signals and for the electronic synthesis of artificial speech. By these means, it became possible to investigate the perceptual relevance of the events observed. The description was

extended with perceptual data.
As a result of these developments in electronics new devices for physiological research were also developed. Knowledge of articulatory postures and movements was increased. With the help of electromyographic measurements the role of individual muscles could be shown.
The introduction of computers during the last decades meant a new impetus for phonetic research. The developments in electronics had already made it possible to build complex models in which various aspects of speech could be combined, e.g. speech synthesizers. However, the computer opened the possibility to simulate even more complex systems and models operating in nearly real time, and to test these quantitatively.

The research presented in this volume mainly concerns the period during which the electronic devices were developed. At the starting point, the investigators worked with recorded speech. The most obvious manipulations that could be performed on recorded speech material were either variations in the time domain or in the frequency domain.

An example of interference in the time domain is the isolation of short stretches of speech sounds (e.g. roughly phoneme or syllable size) from a longer recording. By using these streches in perceptual tests the listeners could better concentrate on the isolated sounds, without being disturbed by unwanted influences from the phonetic context. The isolated samples could also be spliced into various contexts; in this way the perceptual influence of the context on a specific speech sample could be studied (e.g. Fischer-Jørgensen 1969).
The investigations reported upon in this volume belong to a tradition in which the temporal parameters had a central position (Schouten 1967 b); the view taken was that the time element in human perception was underestimated. This line of research had consequences for the model-forming in visual perception (Schouten 1967 a), in pitch perception (e.g. Schouten 1963), in phonetics in general (Cohen, Schouten & `t Hart 1961) and in particular for vowel perception (Cohen et al. 1963) and vowel production (Nooteboom and Slis 1972).
The choice of the time parameter as a lead in speech research did not imply that all other aspects of the speech signal, such as those in the frequency domain, were neglected; it was simply a matter of choice in order to restrict the main objective for research to manageable proportions. In fact, the main objective can be reduced to the question " How much of speech perception can be explained by temporal features ?".
The instrumental aid with which for the experiments on speech in the time domain could be carried out was an electronic gating device, with which it was possible to isolate a stretch of speech with a specific time interval from a tape-loop. The gating function, which determined the amplitude envelope of the isolated sound segment, could be given an amplitude rise time, a total duration and an amplitude decay time at the choice of the experimenter. In this way the amplitude envelope of the stimuli to be judged by the listeners could be kept under control (´t Hart and Cohen 1964).

The alternative way to experiment with prerecorded speech is manipulation and analysis in the frequency domain. An example of this line of research is spectral analysis by means of a sonagraph which yields a spectral display as a function of time, a spectrogram (Potter, Kopp and Green 1947). The majority of the speech scientists concentrated on the frequency domain. The choice of spectral properties of speech as a research objective did not mean, of course, that temporal parameters were totally neglected; these only played a secondary role. The work done in the Haskins Laboratories is an example of the spectral approach.

1.3 ANALYSIS BY SYNTHESIS

Besides experiments with recorded speech, perceptual research was done with pure artificial speech. From perceptual experiments with manipulated speech and from acoustic analyses of the speech signal, hypotheses were formulated with respect to parameters that were important for speech perception. In synthetic speech these parameters could be varied independently of each other. In this way one parameter at a time could be evaluated and thereby "analyzed" perceptually. In this context one speaks of "analysis by synthesis".

In the early speech synthesizers one can find again the dichotomy between the time domain approach and the frequency domain approach.

The main reason advanced for speech synthesis on a segmental basis was based on the finding that, for nearly all phonemes, it proved to be possible to isolate sound segments from a speech utterance that were perceived as "steady states". On this basis it seemed reasonable to assume that an intelligible utterance could be built up from a chain of similar "steady states". The IPOVOX-I was such a speech synthesizer (Cohen 1964). In the IPOVOX-I two sound sources provided the input for a set of resonance filters. The output of these filters could be combined to synthesize segments with one or more formant frequencies. The filtered signals were given adequate amplitude envelopes by function gates for each segment. The IPOVOX-I was a hardware, terminal analog speech synthesizer, which implies that it did not pretend to reflect the human articulatory mechanism, but only to copy the human speech signal. However, human speech is a continuous process. Therefore, in the case of two successive speech sounds with different articulatory configurations, all intermediate positions have to be passed through. This is reflected in the resulting speech sound by gradually changing formant frequencies, the formant transitions. These formant transitions proved to be a powerful cue for the perception of consonants (e.g. Delattre 1962), particularly for stop consonants.
With the IPOVOX-I, gradually changing formant transitions could not be synthesized. Instead, the amplitude decay and rise time of the successive segments overlapped, thus creating a transitional interval during which the spectral information of the first segment was replaced by that of the second segment. In this way the abrupt transition between successive steady states was masked.

Since the steady state part of stop consonants consists mainly of an acoustical silence (or a hardly audible buzzing sound), stop consonants differ mainly with respect to their formant transitions. Gradually changing formant transitions as found in real speech are fundamentally different from the transitional intervals of the IPOVOX-I. It is

therefore not surprising that stop consonants proved to be a problem in speech synthesis based on steady state segments. These problems were the original motivation for the research on voiced and voiceless stops presented here. The investigations extended to related questions concerning voiced and voiceless fricatives and nasals.
The problems with the transitions, in particular with those of the voiced plosives, were solved by introducing an additional "transitional segment", which was in fact a short steady state of about 20-30 ms which faded into the following vowel, giving rise to a voiced stop consonant perception. These short steady state segments were different from real transitions since they did not reflect the continuous movement of the articulators; they acted as it were as "fake transitions" by which the human ear was deceived.

This way of synthesizing proved to be unsatisfactory. Apart from the need to specify a separate transitional segment, which was highly context dependent, the synthesizer produced artefacts at the beginning and end of the segments which were audible as clicks. These originated from the gating devices that defined the amplitude envelopes of the individual segments. The heart of these gates was formed by double pentode tubes. The two sides of these tubes were very difficult to balance, which resulted in clicks especially at points where fast amplitude changes were needed, e.g. in plosives. This was the reason why the IPOVOX-II was constructed. This was also a hardware terminal analog speech synthesizer with two tuneable formant filters by means of which gradual formant transitions could be obtained (Willems 1966). The parameter values to specify formant frequencies, transition durations, amplitudes of voiced and voiceless segments, spectral composition of fricative noise and segment durations could all be controlled by hand.
The procedure of synthesizing speech by hand was laborious and time consuming. Analogous to developments elsewhere in the field of speech synthesis, a set of synthesis rules was defined, which was implemented in a computer. The parameter values that were calculated on the basis of the rules were output in punch tape which served as the input for the IPOVOX-II (Slis 1971 b, Slis & Muller 1971). In writing the synthesis rules, stops provided problems again regarding the optimal values for transitions and formant loci, especially of the second and third formants; questions arose whether one set of loci was sufficient to define one place of articulation, or whether different contexts or different modes of articulation needed different loci.

In the course of time the IPOVOX-II was replaced by a Rockland speech synthesizer with five digital formant filters in series, which was connected on line with a computer (Slis et al. 1977).
The synthesis system with the Rockland synthesizer was followed in its turn by a more flexible fully software synthesis-by-rule system, which will be called the IFN-system (Slis 1978). Both systems operated on a segmental basis. All acoustical segments, of about phoneme size, were defined by 26 parameters, each of which was defined by three values, viz.:

1. a target value specifying the amplitude of one of the sources, a formant frequency or a bandwidth,
2. the duration of the transition interval during which the target

has to be reached,

3. the starting point of the transition relative to the moment of beginning of the sound segment .

In this way the sound segment with its initial transition was fully described. Final transitions were specified as the initial transitions of the next segment. Although the starting points of the various transitions could be chosen independently of each other and a strict segmental control of the synthesizer was therefore not necessary, most rules prescribed in practice the same starting point for all formant transitions.

In the alternative way of speech synthesis special attention was given to spectral properties. The system adopted by the investigators of the Haskins Laboratories operated on a more continuous basis than the IPOVOX-I and II. Their synthesizer was inspired by spectrographic representations. It used hand drawn spectrograms to control the synthesis parameters (pattern playback, e.g. Cooper et al 1952). It was well suited to investigate the timing of various continuous spectral aspects of speech separately, viz. the formant transitions. On the other hand, the IPOVOX-I and its successors, which were based on segmental control of the parameters, were better equipped to investigate the segmental time structure of speech. The difference in equipment defined the difference in approach of the research on speech perception. In the work of the Haskins Laboratories formant transitions took a central position. In several Haskins publications consonants were primarily defined by the formant transitions. Delattre (1962:407) stated that "...les transitions sont à la clef même de la perception de la consonne."

The predominant role ascribed to the formant transitions did not correspond with the experiences obtained with the time domain approach. An experiment was performed in which subjects were asked to identify gated-out portions of speech, viz. the steady state part of a consonant, or its successive vowel including the formant transition of the preceding consonant (Slis & Cohen 1969 b). Identifications of the noise bursts of plosives and of the frictional noises of fricatives were higher (62-91%) than the correctly perceived places of articulation in the formant transitions (22-51%). Besides, it was surprising that it proved to be possible to synthesize intelligible speech with the IPOPVOX-I; a large number of words could be synthesized without using any formant transistions, in particular words containing voiceless obstruents. In some other words short steady states were inserted instead of continuous transitions. These fake transitions proved often to be an adequate replacement for the purpose of word recognition.
Cole and Scott (1967) suggested that the primary function of the formant transitions is to guarantee the continuity of an utterance, in other words that the speech signal will be perceived as consisting of one auditory stream instead of various acoustic streams, which are difficult (or not at all) to relate to each other. In their opinion, too, transitions play a secondary role, in that they add to the perception of the place of articulation of consonants, and most evidently so in voiced stops. It seemed that the primary function of the formant transitions can be taken over by the fake transitions of the IPOVOX-I, in which

continuity is guaranteed by the fading from one segment to the next, thus masking the change in auditory quality. The secondary function, additional information for the place of articulation, which can hardly be missed in voiced plosives, can, to a certain extent, also be taken over by short steady states.
On the basis of data found in the literature and data obtained by his own experiments, Pols (1984 a & b) showed that the relative importance of various perceptual cues is highly dependent on the experimental condition. For example, the perceptual contribution of the noise burst of plosives seems to be significantly larger in isolated VCV stimuli than in the same VCV combinations embedded in sentences.

1.4 THE VOICED-VOICELESS DISTINCTION IN INTERVOCALIC OBSTRUENTS

The reason for this elaborate description of the evolution of methods and tools for phonetic research is that, to a large extent, these define the research objectives chosen. Throughout the research on the voiced-voiceless distinction and assimilation of voice, questions that arose from problems encountered in speech synthesis have been a major motivation. The course taken by the investigations described must therefore be regarded in the light of the development of speech synthesis.
The experiments described in this volume illustrate this evolution. The problems encountered in synthesizing voiced plosives with the IPOVOX-I gave rise to perceptual experiments with manipulated speech and to measurements of the acoustic speech signal. The outcome of these experiments led to the development of a speech synthesizer that could produce formant transitions, the IPOVOX-II. The data obtained with the acoustic measurements and the perceptual experiments with manipulated and with synthesized speech from the IPOVOX-I and II, supplemented with data from the literature are discussed in the first article in this volume (Slis and Cohen 1969 a). At least seven different acoustical parameters proved to be perceptually relevant for the voiced-voiceless distinction. Not all of these parameters are independent of each other. In the second article an attempt was made to relate these parameters to each other (Slis and Cohen 1969 b). For some of the parameters mutual dependency was evident. Other parameters, however, could only be related to each other on the basis of assumed common articulatory events. The measurements on the voiced-voiceless distinction we found in the literature were made for languages other than Dutch; descriptions based on them may not be applicable to this language. The collection of data from acoustical and perceptual experiments in Dutch and other languages and articulatory data from foreign languages only were fused into one articulatory/acoustic framework which represented the differences in the production of voiced and voiceless obstruents.
Since the acoustic and the articulatory features of voiced and voiceless obstruents were reported to differ in various languages, the need was felt to check some of the articulatory data with measurements on Dutch speakers. These measurements were the subject of the third article (Slis 1970) in this volume. With the help of the new data the articulatory model of the voiced-voiceless distinction was further developed. One of the aims in the design of the revised model was to reduce the large number of parameters that were found to accompany the voiced-voiceless

distinction to one single cause, in order to adapt these to the linguistic notion of only one distinctive feature. To this end a feature "articulatory effort" was suggested.
In a fourth article (Slis 1971 a) an attempt was made to also relate the model of the voiced-voiceless distinction to other linguistic oppositions in which articulatory effort played a role. A model was designed in which one feature "articulatory effort" was assumed to be the ultimate cause of the voiced-voiceless distinction as well as the difference between long and short vowels and between stressed and unstressed syllables. An important element of this model is the hypothesis that gestures executed with increased effort are advanced in time.

The model that emerged after the first four publications was based on a difference in articulatory effort of all muscles involved. The consequence for voiceless obstruents compared to voiced ones can be summarized in the following way (the acoustic consequences are identified by (1) to (8)):
A) An increase in the activity of the muscles that are responsible for the oral constriction leads to an advancement in time of the closing command and to an increase in speed of the gesture. This results in an earlier moment of oral closure; consequently (1) the vowel duration of the preceding vowel will be shortened and (2) the duration of the consonant constriction will be lengthened. (3) The formant transitions of the preceding vowel will be shortened because of the faster closing gesture.
B) An increase in the activity of the pharynx musculature results in a contraction of the pharynx cavity and an upward movement of the larynx. As a consequence of both events the volume of the pharynx will be decreased. This can be regarded as a pumping action as a result of which the intra-oral air-pressure will rise.
C) The mechanisms that change the larynx position and the changed position itself lead to less favourable vibration conditions of the vocal cords. It is assumed that they take a slightly abducted position and that tension is increased. Consequently, (4) vocal cord vibration will be interrupted during voiceless obstruents. In addition, the tenseness of the vocal cords leads to (5) a higher pitch and (6) a lower amplitude in the adjoining vowels. Also the resumption of vocal cord vibration will be delayed, leading to a relatively (7) long voiceless (noisy/aspirated) interval between the moment of oral opening and voice onset time (VOT). The voiced part of the (3) initial formant transitions will therefore be shortened since a large proportion of the opening gesture takes place before the VOT.
D) As a result of a lower resistance in the slightly opened glottis for the airstream from the lungs to the pharynx, the difference between the supra-glottal pressure in the pharynx and the subglottal pressure in the trachea will decrease quickly. In combination with the pumping action of the pharynx, the intra-oral pressure will reach relatively high values. In fricatives this will result in (8) a strong friction noise that originates in the constriction opening; for the same reason the burst at the release of plosives will be loud.

After the publication of the first four articles presented in this volume additional articles appeared in the literature on this subject. In an inserted chapter some of these later articles will be discussed in order to update the model proposed. In particular the mechanism that was held responsible for the interruption of vocal vibration, mentioned under (C) above, had to be revised. New evidence showed clearly that in addition to a difference in effort a different programming of the larynx musculature had to be assumed.

1.5 COARTICULATION OF EFFORT IN CLUSTERS

The results of the study of intervocalic obstruents have a limited value of course. A wider application of the model can only be claimed after the voiced-voiceless distinction has been studied in other contexts. Therefore, experiments were done with intervocalic two-consonant combinations, which will be referred to as clusters in this volume. The consonants were part of two different syllables (V1C1-C2V2); C1 is the final post-vowel consonant of the first syllable and C2 the initial pre-vowel consonant of the second syllable. Measurements of oral opening and closing showed that the articulations of C1 and C2 overlapped to a great extent. As a consequence, the realisations of the two successive syllables overlap also. From the results of a first experiment it was concluded (Slis 1972, Slis 1975) that articulatory qualities of the second syllable were already executed in gestures belonging to the first syllable; the vowel of the first syllable was shortened because of an advancement of the moment of closure of C1 if C2 (of the subsequent syllable) was given extra effort because of stress. It was hypothesized that the effort that had to be applied at a certain time for one set of articulatory commands, belonging to one syllable or phoneme realisation, was also administered to other, simultaneously executed gestures. This can be viewed as a coarticulatory effect.
A large set of measurements on two-consonant clusters, in which C1 and C2 could be voiced or voiceless plosives, or nasals, was performed in order to specify timing rules for speech synthesis (Slis et al. 1977). A set of seven rules for the timing of two-consonant clusters, consisting of all combinations of voiced and voiceless plosives and nasals, were obtained and implemented in a synthesis by rule system. The application of these rules yielded a timing pattern which was in accordance with measurements of intervocalic obstruents.

We can summarize these seven cluster rules by the following three statements:
1) The presence of effort of a voiceless C2 is regressively assimilated, which results in advancement of the oral closing gesture in a preceding nasal.
2) Repetition of a feature of a C1 in the C2 of a C1C2-cluster results in shortening of the C2.
3) Rules for single intervocalic plosives and nasals are also valid in two-consonant clusters.

In a subsequent paper (Slis 1979) some measurements of the time structure of three-consonant clusters were reported in which the second consonant was /s/. Comparison of the clusters /psp/, /msp/ and /tsp/ with /psb/, /msb/ and /tsb/ showed again that /p/ closure duration is longer than that of /b/. As was the case in two-consonant clusters the difference in duration of cluster final /p/ and /b/ cannot be due to an advancement of the time of closing the oral tract of /p/ at the expense of the preceding /s/, since /s/ duration was equal before /p/ and /b/. Again the difference must be due to shortening of /b/ compared to /p/.

One restriction has to be made. The measurements of two- and three-consonant clusters were done in order to find rules for speech synthesis. It was suggested that timing rules obtained from only one speaker of acceptable Dutch, should yield at least one set of acceptable time structures for synthetic speech. Since the findings thus obtained from one speaker confirm earlier data as well as the findings in the literature, it seems probable that these rules are representative of Dutch.

Although the underlying processes were not fully understood, a good view of the voiced-voiceless distinction was obtained and a satisfactory model giving an adequate description of the articulatory events accompanying this distinction was created. The model served as a basis on which the timing rules for speech synthesis were based. A missing component in these rules concerned the timing of voice-activity during consonant clusters. Therefore, this aspect in particular was studied in subsequent research. These voicing problems are traditionally part of the topic of our research, viz. assimilation.

1.6 ASSIMILATION OF VOICE

In spite of the fact that assimilation of voice has been studied extensively in classical phonetics, and that a rich recent literature is present on the voiced-voiceless distinction and on coarticulation in general, surprisingly little recent research has been published on assimilation of voice from a phonetic point of view; the 5th, 6th and 7th article in this volume (Slis 1983 a, 1983 b, 1985) deal with assimilation of voice.

The approach to the study of the voiced-voiceless distinction and that to assimilation of voice were quite different. In the processing of the data of voiced and voiceless consonants, all measurements of the phoneme realisations were considered to be part of one population; without regarding effects of assimilation, average values were calculated. On the other hand, in the available relatively old literature on assimilation, the data were presented in terms of frequencies of progressive, regressive or no assimilation separately. Almost no quantitative data on measurable aspects of the acoustic signal were given.
A start was made to fill the gap between the observations of the classical literature in terms of frequencies of assimilation and the timing data obtained in the experiments on the coarticulation of effort. Besides, it was attempted to find the roots of the inconsistencies in the data that were presented in the literature. In order to obtain data that could be compared with those of the literature, definitions had to be formulated of the criteria by which the recordings of clusters could be divided into the three usual assimilation classes on a quatitative basis.

The definition of assimilation of voice was based on the one given by Crystal (1980:35) who says that assimilation is " A general term in PHONETICS which refers to the influence exercised by one sound segment upon the ARTICULATION of another, so that the sounds become more alike, or identical". If articulatory realisations could be measured that differed significantly from what is normal for voiced and voiceless consonants, assimilation occurs. For that purpose timing of voice offset and onset relative to the moments of opening and closing of the vocal tract was used (Slis 1982 a, 1983 a (5th article in this volume), 1985 (6th article in this volume)). These moments were derived from simultaneously recorded oscillograms of the acoustic speech signal and of voice activity measured by means of an electro-glottograph. A sudden decrease of the speech signal indicated the moment of closing of the oral tract, and the presence of a noise burst or a sudden increase of the speech signal indicated the moment of oral opening. Moments of voice offset and onset were found in the electrolaryngogram at points where the signal ceased and started again.
Criteria for voice offset and onset times were obtained from measurements of intervocalic stops. In voiceless plosives, voice offset occurred 20-30 ms after oral closing (Slis 1970). The time interval between oral closing and voice offset time was called "voice tail". The standard deviation in the duration of the voice tails proved to be about 10 ms. On this basis, originally voiceless post-vowel obstruents (C1) in two-consonant clusters which have voice tails longer than 50 ms, were defined operationally to be voiced obstruents by assimilation. The chance of an erroneous decision is less than 2.5 % with the 50 ms criterion, viz. the mean voice tail duration plus two times its standard deviation. In intervocalic voiced stops, voice activity continues uninterruptedly during the consonantal interval; in initial voiced plosives the voice onset takes place about 25 ms before the moment of oral opening. Analogous to the reasoning for voice offset, originally voiced pre-vowel obstruents (C2) in two-consonant clusters with a VOT equal to or larger than zero, were operationally defined to be voiceless obstruents by assimilation.
Assimilation of voice was studied in two-consonant clusters (C1-C2) consisting of a voiceless syllable-final obstruent (C1) followed by a voiced syllable-initial obstruent (C2). With these clusters all cases are covered where assimilation between obstruents in C1C2 clusters in Dutch may occur, since the obstruent C1 is always voiceless because of a final devoicing rule in Dutch (neutralisation). Assimilation of voice can only occur if the two successive consonants originally differ with respect to their voice character. Consequently, as C1 is always voiceless, C2 must be voiced.

1.7 INFLUENCES ON ASSIMILATION OF VOICE

From the literature, which stems mainly from before the 1960's and which is therefore based on subjective data, two general assimilation rules can be formulated. These two rules are accepted by Dutch phonologists. They are:

(1) regressive assimilation is observed in the majority of two-consonant clusters if the second consonant is a voiced plosive;
(2) progressive assimilation is generally found if the second consonant

is a voiced fricative (Slis 1982 a, 1985 (6th article in this volume)).

Consensus among the various authors is found with respect to the opinion that a number of deviations from these rules have to be assumed. Sixteen different causes that give rise to deviations were mentioned in the literature (Slis 1982 a, 1985).

Investigations of eleven of these factors that influence assimilation were executed, viz:

1) Pitch variations: speech with normal intonation vs. monotonous speech.
2) Pitch height: low, medium and high pitched monotonous speech.
3) Voice quality of the speakers: good vs. poor quality.
4) Sex of the speakers: male vs. female speech.
5) Speech rate: slow, normal and fast speech.
6) Phonological composition of the clusters with respect to manner of articulation: stop-stop, fricative-stop, stop-fricative and fricative-fricative clusters.
7) Phonological composition of the clusters with respect to place of articulation: /p,t,k,f,s,x/ followed by /b/ or /d/.
8) Stress conditions: before or after a stressed vowel, or without stress on the adjacent syllables.
9) Phonological length of the preceding vowel: /a:,o:,ø:,e:/ vs. /ɑ,ɔ,æ,ɛ,I/.
10) Linguistic boundaries: across syllable boundaries within a word vs. across word boundaries.
11) Region of origin of the speakers: dialectal background.

The results of the experiments regarding points 1 up to 10 are summarized in Slis (1984). Results on points 4, 5, 6, 7, and 11 are given in more detail in Slis (1982 a, 1985 (6th article in this volume)), on points 1, 2, 3 and 4 in Slis (1983 a (5th article in this volume)) and on points 8 and 10 in Slis (1983 b (7th article in this volume)).

1.8 ASSIMILATION REGARDED AS COARTICULATION

The working hypothesis for the experiments on assimilation of voice was that assimilation is a special form of coarticulation. The roots for this hypothesis were already detectable in the former work on the timing in two-consonant clusters (see above). If two successive consonants (overlapping in time) have to be articulated, it may occur that during the period of overlap two conflicting gestures have to be executed. This is the case in C1-C2 clusters in which C1 is a voiceless and C2 a voiced obstruent. In cases where two parties have conflicting interests, there are two ways to solve the problems.

First, one of the parties withdraws its claims. In the context of assimilation this means that assimilation belongs to the domain of the motor program to the articulators; in this case articulation is preprogrammed on a high neurological level.

Second, the two parties fight and the strongest wins; this is a case of coarticulation in which peripheral mechanisms define the ultimate realisation.

The second solution is favoured here for a number of reasons:

1) If assimilation were preprogrammed, one would, on the basis of economy, expect consistent assimilation behaviour, e.g. a certain speaker should always use regressive assimilation in a certain context under fixed conditions. This proves not to be the case. The direction of assimilation is not easily predictable (Slis 1982 a, 1985).
2) The factors investigated often influence assimilation to a different degree. This implies that each factor needs its own extra rule to calculate the direction and the degree of assimilation. This presupposes a complicated system of rules, which is superfluous if we assume that assimilation is a peripheral consequence of mechanical coarticulation.
3) The differences in assimilation that are due to the experimental variables can be explained by mechanical parameters of the articulators, which should influence voicing anyway. It seems a waste to introduce pre-programmed motor commands on top of that.

Let us consider which conditions influence vocal cord vibration.
(A) First of all, glottal width is an important parameter in this respect. As argued with the model of the voiced-voiceless distinction, a larger separation of the vocal folds (as in voiceless consonants) results in a lower glottal resistance. Consequently, the pressures above and below the glottis are equalized during a short time. If the pressure drop across the glottis is too small, glottal vibration will cease. Besides, a wider glottal aperture will lead to a slower airstream velocity through the glottis, and consequently to less underpressure between the vocal folds (smaller Bernoulli effect).
(B) Secondly, the tension in the vocal cords influences vocal cord vibration. This becomes manifest as a higher pitch with more tension, as e.g. in vowels adjacent to voiceless consonants. Besides, Halle and Stevens (1971) calculated that with stiff vocal cords a larger pressure drop across the glottis is necessary to maintain vibration than with slack ones.
(C) A third point to consider is the vibrating mass of the vocal cords; a larger mass leads to lower, frequencies as e.g. in male larynges compared to female ones. Besides, a larger vibrating mass is often accompanied by less tense vocal folds. These two variables, small tension and large mass, induce a favourable condition to keep the vocal cords in vibration.
(D) In the fourth place, the size of the supra-glottal cavity determines the amount of air that has to be transported through the glottis before the pressure drop diminishes to an extent that vocal fold vibration stops. This aspect is also discussed in the model describing the voiced-voiceless distinction.
(E) Finally, a fifth parameter that may cause an interruption of voicing is the duration of the closure of the vocal tract. If this duration is sufficiently long, the supra-glottal pressure will become equal to the subglottal pressure, and consequently vocal vibration will cease.

The results of the experiments on the eleven factors mentioned earlier that may influence assimilation of voice can all be interpreted in terms of these five parameters. In the epilogue this is elaborated for the eleven factors separately. The result can be viewed as a confirmation of the hypothesis that assimilation of voice can be regarded as a special case of coarticulation.

CHAPTER 2

ON THE COMPLEX REGULATING THE VOICED-VOICELESS DISTINCTION I

I. H. Slis and A. Cohen

Reprinted from:

LANGUAGE AND SPEECH
Vol. 12, part 2, 1969, 80-102

ON THE COMPLEX REGULATING THE VOICED-VOICELESS DISTINCTION I*

I. H. SLIS and A. COHEN
Institute for Perception Research, Eindhoven

An investigation is described of the perceptual and acoustical differences between voiceless and voiced consonants in Dutch. The experiments reported included synthesized, normal and whispered speech. The results are compared with those found in the literature for other languages. Parallels between the results of the various investigations could be established, although quantitative differences apply to different languages.

It is shown that a number of perceptual and acoustical cues are present, such as: sound intensity and duration of fricative noise, duration of the consonant and duration of the preceding vowel, voicing and the moment of voice onset, intonation pattern and amplitude as well as duration and frequency range of the formant transitions of the adjoining vowels.

To establish a connection between these parameters an articulatory model of the difference voiced-voiceless is proposed. Essential features in the construction of this model are the observed acoustical symptoms and physiological data as reported in the literature. The model is based on the assumption that the difference voiced-voiceless is due to the presence or absence of activity of the pharyngeal constrictor muscle.

This muscle influences the volume of the pharynx and the position of the larynx, and consequently the pressure drop across the glottis and the vibration condition of the vocal cords.

From recent literature it is evident that the voiced-voiceless distinction is still a burning issue in phonetics research. In linguistic analysis the description of this distinction in terms of a single label seems to suffice. By this means plosive and fricative consonant phonemes can be easily systematised as voiced (e.g./b,d,g/ and /v,z/) *vs.* voiceless (p,t,k/ and /f,s/). Inspired by this linguistic notion of a single feature on which this distinction can be based, phoneticians have been at pains to lay hold of a concrete correlate of this linguistic feature in studying articulatory, acoustic, and perceptual phenomena of speech. Surveying the literature one is forced to the conclusion that in all three domains the voiced-voiceless distinction is, in fact, carried by a number of attributes depending on the language and the speech context in which it is studied.

In the present article an account will be given (a) of the various attributes to be found in the literature and (b) of those which can be established as applying to the voiced-voiceless distinction in Dutch, as measured acoustically with certain specified contextual constraints in actual speech; (c) of the extent to which the most

* *The second part of this paper will appear in the following issue of* Language and Speech.

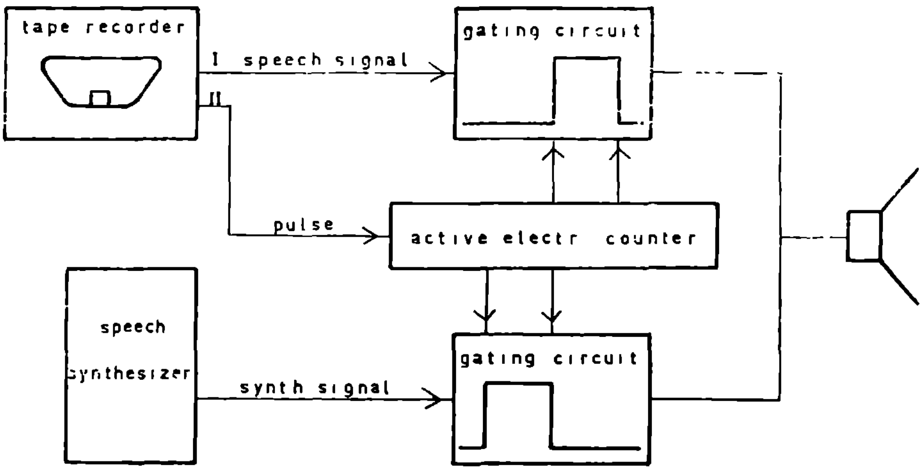


Fig. 1. Block diagram of the set-up for carrying out commutation of real with synthetic speech.

prominent attributes determined acoustically can be shown to have perceptual consequences, including the use of synthetic speech; (d) of a tentative model based on articulatory data derived from literature to explain the relationship of the various attributes contributing towards the voiced-voiceless distinction in Dutch, possibly with a wider application (see also Slis, 1966, 1967).

Apparatus and procedure

The acoustic measurements were carried out on words of the types /bəCœt/, /baCə/ and /bɑːCə/ (in which /C/ can stand for /p, t, b, d, f, s, v, z/)

These words were normally spoken and whispered by two male speakers in a sound treated booth and recorded on a semi-professional magnetic tape recorder, 3.75 in/sec., type Revox nr. G 36. The recorded signal was registered by a UV oscillograph, (Visicorder, Honeywell type 906S) to determine duration and amplitude. To measure fundamental frequency (F_0) as well, use was made of a pitch meter designed at the I.P.O.

In perceptual tests involving real speech, use was made of an electronic gating circuit for isolating segments of any required length from utterances recorded on a loop of magnetic tape (see 't Hart and Cohen, 1964).

In tests with synthetic speech the signals used were generated by either of two terminal analogue speech synthesizers, designed at the I.P.O. The older version, IPOVOX I, was based on the principle of cascading steady-state segments of phoneme size obtained by filtering with single tuned resonance filters and by modulating the amplitude by electronic gating circuits (Cohen and 't Hart, 1962a; Cohen, 1964).

In the more recent version, IPOVOX II, use was made of two electronic variable filters through which it is possible to introduce gradual glides in connecting the segments (Willems, 1966).

In tests in which both real and synthetic speech were used, concrete commutation was carried out with a set-up capable of exchanging part of a spoken utterance recorded on a tape loop by a synthetic signal. These two signals are defined in time by two electronic gating circuits which are triggered by an active electronic counter which itself is triggered by a pulse recorded on the second track of the tape loop, see Fig. 1.

In the perceptual tests 3 - 6 subjects were used, all native speakers of Dutch and, as employees of I.P.O., accustomed to hearing synthetic speech.[1]

In all perceptual tests the subjects were asked to identify the sounds they heard as speech sounds. They were allowed free choice in their responses although they were aware of the fact that the tests were meant to obtain information on voiced and voiceless plosives and fricatives.

Whenever synthetic speech was included the subjects were asked to indicate the stimuli they considered as of exceptionally good quality. In scoring the results such stimuli were given two marks compared with one mark for plain identification and zero for ambiguous answers.

The outcome of the experiments will be represented as percentages of the marks obtained for each stimulus.

Acoustic attributes of the voiced-voiceless distinction

In the following sections, attributes of the consonant proper will be dealt with first. As the presence or absence of voice is the oldest attribute put forward to distinguish the consonants at issue, it seems natural to start with it. Next, friction noise will be discussed, being the acoustic counterpart of the periodic voice attribute.

In the following two sub-sections the contextual repercussions of the presence or absence of voice will be taken into account, first with regard mainly to the length of the preceding vowel, and subsequently to the formant transitions and the fundamental frequency and intensity of adjoining vowels.

In each section the plosives will be dealt with first and then the fricatives unless both groups can be shown to share the attribute at issue.

[1] *Since it is our belief that " . . . users of a particular language have a built-in pattern which enables them to identify certain acoustic signals as sounds of their language and to reject others . . ." (Cohen et al., 1963), we thought it justified to consider the judgments of a few listeners as representative of the conduct of Dutch language users in general. We are strengthened in this opinion by an experiment on vowel perception employing a big panel of listeners (Cohen et al., 1967) which showed no significant differences from an earlier experiment with only a few listeners.*

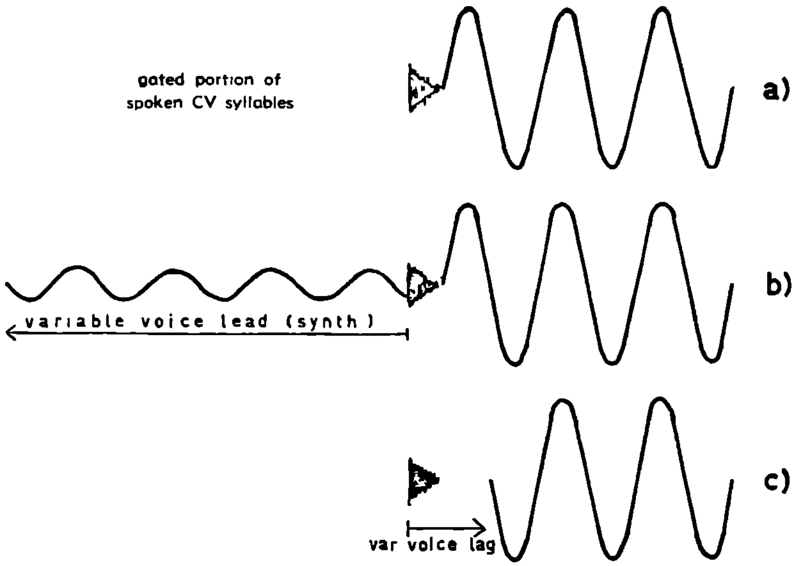


Fig. 2. Stylized oscillographic display of the stimuli with voice lead and voice lag.
(a) The original spoken CV syllable.
(b) The same with a synthetic voice lead added.
(c) The same with introduction of a silent interval to create the effect of a voice lag.

Voice

Lisker and Abramson (1964) measured the moment of voice onset with plosives relative to the release of the stop in 11 different languages including Dutch. It turned out that in all investigated cases in CV contexts the moment of voice onset with voiced plosives occurred earlier than with voiceless plosives. This means, in effect, that in most of these languages in voiced plosives this moment occurs before the release of the stop; in other words, voice is present during the consonant itself, in contradistinction to voiceless plosives where the voice starts after the release.

In another article, the same authors (Abramson and Lisker, 1965) report on a perceptual experiment with English, Spanish, and Thai-speaking subjects. They were asked to identify synthetic stimuli differing only in the moment of voice onset. The subjects proved to be able to distinguish on account of this single cue between voiced and voiceless plosives and the numerical values established correlated well with those found in their preceding acoustical measurements with real speech (cf. Fujimura, 1961 and Lisker and Abramson, 1965).

To find out the perceptual relevance of the acoustic attribute for Dutch inherent in the voice onset time, a commutation test was carried out. For this purpose we used the experimental set-up mentioned above, see Fig. 1. The spoken syllables /ta, da, sa, za/ were recorded on the tape loop. Of those syllables the vowel part and the last 10 msec. of the consonant part were isolated. In this way 4 more or less similar segments were obtained made up of about 10 msec. noise, a vowel formant transition, and the vowel /a/. Next, a periodic segment of variable length was synthesized consisting of a small number of harmonics around 200 c.p.s. with a fundamental frequency of 100 c.p.s. and of a level well below that of the consonant part. Thus a voice lead ranging in six steps from 200 to 40 msec. could be produced.

To obtain voice lag, a soundless interval was substituted for the part of the vowel immediately behind the 10 msec. consonant part, which was varied from 0 to 30 msec. in steps of 10 msec. making the effective voice lag vary from 10 to 40 msec., see Fig. 2.

The stimuli were randomised and presented to three subjects. The results presented in Fig. 3a that voiced judgments correlate well with a voice lead and voiceless ones with a voice lag. The top (Fig. 3b) representing the number of 'excellent' judgments indicated an optimum value for voiced judgments with a voice lead of 80 msec. This value fully agrees with the acoustical measurements of Lisker and Abramson (1964) for Dutch. A similar, though less definite peak could be established for voiceless judgments. However, the peak value found perceptually (10 msec., Fig. 3b) could hardly be expected to correlate well with the 20 msec. measured in real speech, since the introduction of a gap within the stimuli with voice-onset times longer than 10 msec. caused them to sound less natural.

In fact, this gap gave rise to erroneous judgments with regard to the place of articulation (*viz.* 10% /p/ judgments).

In the stimuli with a voice lag, in a number of cases voiced judgments were made which turned out to derive from stimuli that were originally spoken syllables containing a voiced consonant (about 30% of the /da/ and /za/ stimuli). This seems to indicate a possible influence of attributes other than the moment of voice onset in distinguishing voiced from voiceless plosives.

With the fricatives it seems more relevant to deal with the presence of voice (e.g. Jassem, 1962) rather than the moment of voice onset. In addition to a clearly periodic part of the spectrum, the friction noise could be shown to manifest itself as modulated by a frequency corresponding to F_0 (Tscheschner, 1966).

Similar observations from spectrographic and oscillographic recordings could be obtained for Dutch.

Results of perceptual tests by Forrez (1966) carried out at the I.P.O. with synthetic stimuli showed that an improvement of /z/ perception can indeed be brought about by merely adding voice to the friction noise and further improvement could be made by modulating the noise by the fundamental frequency.

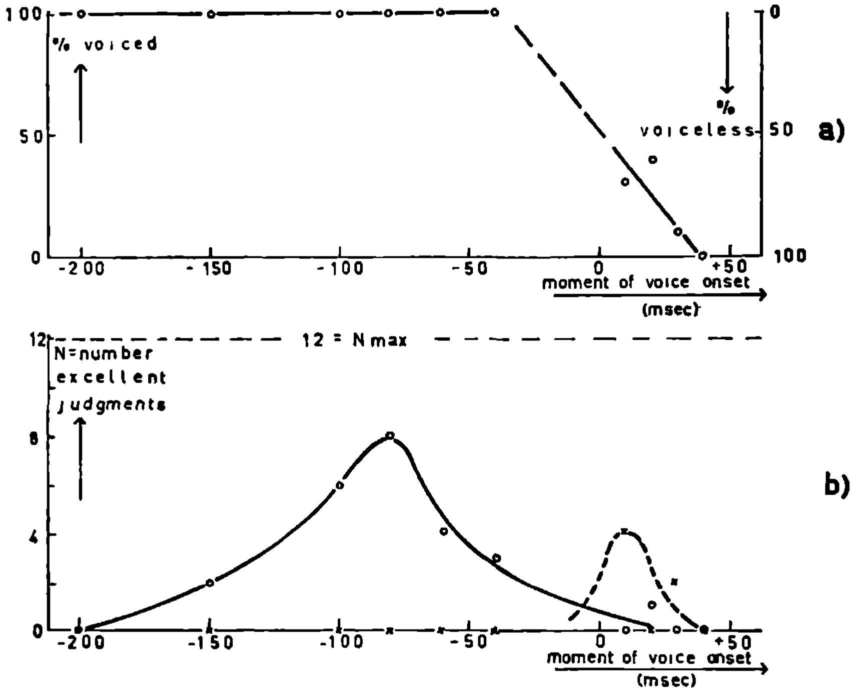


Fig. 3. (a) Voice character as a function of moment of voice onset as obtained in a perceptual test with CV stimuli. (Each point in the graph has been obtained from 12 judgments.)
(b) The number of "excellent judgments" for voiced and voiceless responses respectively as derived from the same measurements as illustrated in (a).

Friction Noise

The friction noise can be described in terms of its duration, intensity and spectral composition. First duration and intensity of plosives and fricatives will be tackled separately, next the spectral composition of both groups together, as homorganic plosives and fricatives are considered to have close spectral affinity. With regard to plosives, duration measurements can be made both of the noise burst proper and the preceding so-called silent interval (also referred to as occlusion).

Plosives

It is generally acknowledged in the literature that the noise burst of voiceless plosives has a longer duration than that of its voiced counterpart (e.g. Fischer-Jørgensen, 1954, Vieregge, 1966).

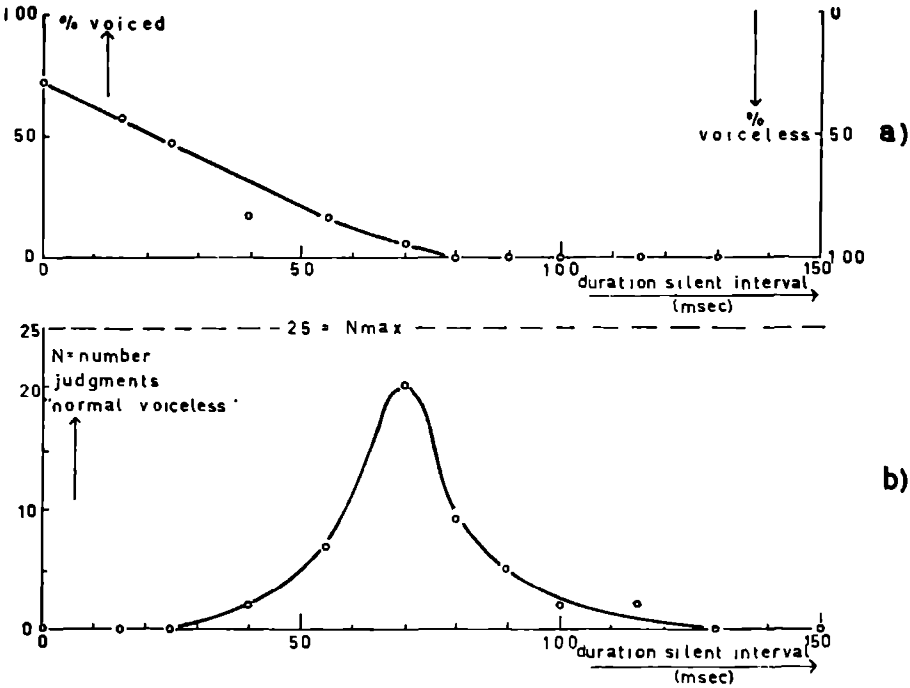


Fig. 4. (a) Voice character as a function of the duration of the silent interval as obtained in a perceptual test with words of the type /lɑCə/. (Each point has been obtained from 25 judgments.)
(b) The number of " excellent judgments " for voiceless plosives as derived from the same measurements as illustrated in (a).

Although no specific measurements of the silent interval seem to be available in the literature, tests have been performed of the perceptual contribution of this factor. For English this was studied with the synthetic words *rapid-rabid* where it was shown the longer interval tended to favour voiceless and the shorter voiced judgments (Liberman *et al.*, 1961).

The noise burst duration of the plosives in spoken Dutch words measured on an oscillographic display was found to be 15 msec. (standard deviation of the mean, $s = 4.5$ msec.) longer for voiceless plosives than for voiced ones in otherwise identical circumstances. Similarly, the duration of the silent interval proved to be 28 msec. ($s = 5$ msec.) longer in the case of the voiceless plosives. No significant difference from these values was found when these words were whispered.

Although specific perceptual tests were of no avail[1], our general experience in synthesizing words correlates well with the acoustical data that longer noise bursts favour a voiceless impression.

To establish the effect of the silent interval in Dutch words of the type /lɑCə/ were synthesized, in which /C/ was made up of a silent interval which could be varied from 0 to 130 msec. and a noise burst of constant duration (10 msec.). The spectral composition of the burst was chosen to correspond with three different places of articulation.

The results as represented in Fig. 4 show that short silent intervals give rise to a voiced sensation, and longer intervals to a voiceless one.

Just as a longer duration of the noise burst was acknowledged to be characteristic of voiceless plosives, so can a similar observation with respect to a higher intensity of the noise be found in the literature (e.g. Halle *et al.*, 1957; Strevens, 1960). For Dutch we also obtained measurements of the amplitude of the noise burst of plosives in a way similar to that described above. The amplitude of the voiceless burst was about 50% (s = 16%) higher than the amplitude of the voiced one. Again no significant difference from these values could be found in whispered speech.

As for the perceptual consequences[2], here again our general experience in synthesizing words correlates well with the acoustical measurements, *viz.* that a higher noise intensity favours voiceless plosive judgments.

The perceptual tests carried out at an early stage in this investigation to determine the influence of intensity and duration of the noise burst of synthetic plosives all gave rise to voiceless judgments only. This outcome may have been due to too high a noise intensity. The same factor may also have played a part in subsequent tests resulting e.g. in the comparatively large number of voiceless judgments in our experiment on the influence of the duration of the silent interval mentioned above (cf. Fig. 4).

However, a differentiation between labial and velar plosive judgments offered itself on account of differences in amount of noise only (duration and intensity), /k/ sensation corresponding to large and /p/ sensation to small values, see table below.

Duration of burst	% Judgments /p/	% Judgments /k/	Sound Level of Burst rel. to Vowel	% Judgments /p/	% Judgments /k/
10 msec.	58	42	−18 db	62	38
15 msec.	31	69	−13 db	57	43
25 msec.	27	73	− 8 db	4	96

Similar correspondence between place of articulation and amount of noise are testified to in the literature on the basis of acoustic measurements both for duration (e.g. Fischer-Jørgensen, 1964) and intensity (e.g. Halle et al., 1957).

Fricatives

As with plosives throughout the literature evidence is found of a coupling of higher noise values (duration and intensity) with voiceless fricatives (e.g. Strevens, 1960). Apart from the acoustical measurements, perceptual consequences of this attribute are reported, notably with respect to the duration of the friction noise (Denes, 1955); by reducing the length of the friction noise of /s/ in the word *use* (noun), listeners could be induced to identify it as *use* (verb).

For Dutch we found both for normal and whispered speech the mean difference in length between voiced /z, v/ and voiceless /s, f/ to be 50 msec. ($s = 5$ msec.) and in amplitude 25% ($s = 7\%$).

Spectral composition of friction noise

For the generation of acceptable synthetic fricatives, it was necessary to determine the spectral composition of the various fricatives to be used (/s, f, x/). Peaks in the noise spectrum, as established through perceptual tests to lead to high identification scores, are plotted against the second formant frequency F_2 of the following vowel for /s, f, x/ in Fig. 5. These values agree reasonably well with those found in the literature (e.g. Strevens, 1960).

Work done at the Haskins Laboratories (Cooper *et al.*, 1952) shows that it is possible to identify plosives by their noise spectrum. Noise with a spectral peak at a high frequency (2500 c.p.s. and higher) resulted in /t/ responses, with a spectral peak lower than 2500 c.p.s. resulted in /p/ or /k/ responses; the choice /p/ - /k/ was dependent on the vowel used and on the frequency of the spectral peak of the noise burst.

Acoustic measurements indicate that a similar noise quality is present in plosives and fricatives of the same place of articulation (e.g. Fischer-Jørgensen, 1954). Moreover, when the initial part of the spoken fricatives is suppressed by means of an electronic gating circuit, leaving the final 10 - 30 msec. intact, the corresponding plosives are perceived (Cohen and 't Hart, 1962b). On the basis of these acoustical and perceptual findings we felt justified in using the same spectral composition for synthetic homorganic plosives and fricatives.

Preceding Vowel

Numerous investigators have testified to the fact that vowels preceding voiced consonants are generally longer than those preceding voiceless ones, e.g. for American English, House and Fairbanks, 1953, Belasco, 1953, Peterson and Lehiste, 1960 and for whispered speech, Sharf, 1964; for Spanish, Zimmerman and Sapon, 1958; for French, Delattre, 1962b.

The perceptual consequences of the duration of the preceding vowel are established quantitatively by Denes (1955) with the synthetic word *use* in the article referred to above. With the long duration of /u/ the following consonant was identified in the majority of cases with /z/, with short /u/ values, with /s/.

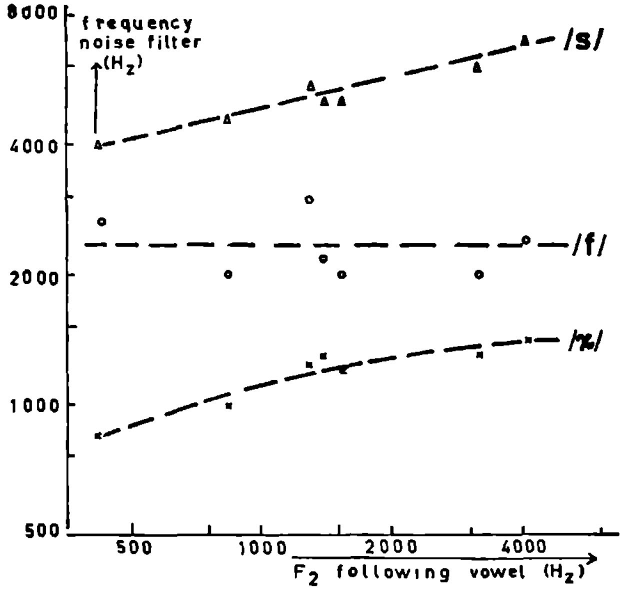


Fig. 5. Peaks in the noise spectrum of the fricatives /s,f,x/ as a function of F_2 of the following vowel, as obtained in a listening test with synthetic stimuli.

Our own acoustical measurements show that if the consonants differ in voice character only, a particular vowel preceding a voiced consonant is always longer than when followed by a voiceless consonant within the same context. The mean of this difference is 30 msec. (s = 3 msec.) with plosives and 40 msec. (s = 3 msec.) with fricatives. This applies both to normally spoken and to whispered speech.

In a perceptual test in which, unlike the other experiments with synthetic speech, the subjects were asked to adjust the vowel length in words of the type /hVCə/ (in which V stands for /a, ɛ, ɪ, e, i, ɛi/ and C for /f, s, p, k, v, z, b, d/). The object of this test was to find out whether a perceptual influence on the length of the vowel could be established depending on the voice character of the following consonant. Vowels preceding voiced consonants turned out to be 25 msec. (s = 2 msec.) longer than those preceding voiceless ones. No difference was observed with regard to plosives or fricatives.

Transitions

Vowel formant transitions (to be referred to as 'transitions') manifest themselves as frequency shifts in spectrographic recordings of the vowel formants at the moment in time when consonant and vowel meet. We shall restrict ourselves to plosives as no

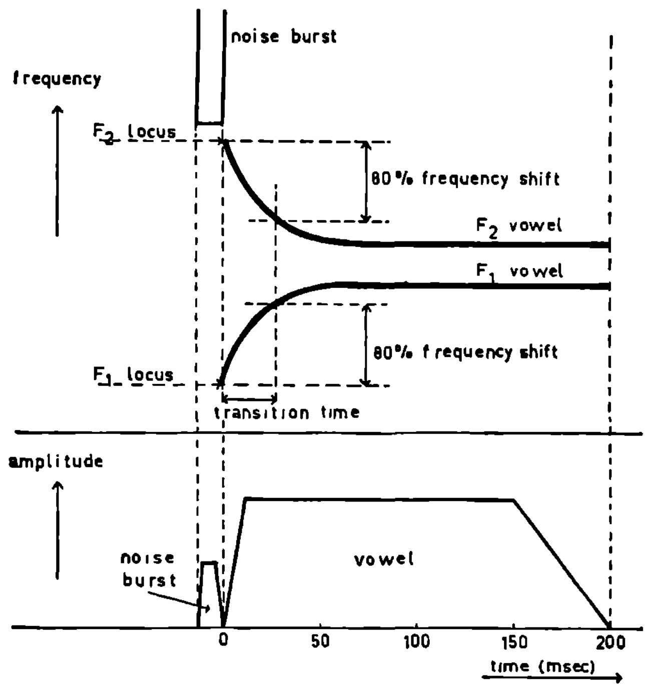


Fig. 6. Schematic view of the synthesized CV stimuli. In the top part a definition is given of locus and vowel transition time. (Notice transition time of F_1 and F_2 is the same.) The amplitude envelope is shown in the bottom part.

information in this respect seems to be available with regard to fricatives on the voiced-voiceless issue.

Spectrographic studies (Halle *et al.*, 1957, Lehiste and Peterson, 1961a) seem to show that it is very hard to obtain information on the influence of transitions on the voiced-voiceless distinction. In the words of Halle *et al.* (1957) " Our first problem, how to define a transition, illustrates well the difficulties that are encountered in the study of natural speech, but can easily be avoided when one has control over the stimulus " (p. 112). Therefore, it seems rewarding to look at the results provided by perceptual tests. Especially the work done in the Haskins Laboratories should be mentioned .

To start with, the speed of the transitions is generally taken in their work to provide a cue for the voiced-voiceless distinction: rapid transitions tend to voiceless and slow transitions to voiced plosive judgments (e.g. Delattre, 1962a). (Extremely slow transitions gave rise to the perception of the glides /w, j/, Liberman *et al.*, 1956.)

Furthermore, great attention is paid to the part played by the transitions of the first formant F_1, notably the range of the frequency shift: a small range causes voiceless, and a wide range voiced plosive judgments (Cooper *et al.*, 1952). A further cue in connection with the F_1 transition has been investigated by delaying the onset of F_1 relative to F_2: a longer delay makes for a voiceless impression (Liberman *et al.*, 1958).

The major cue inherent in formant transitions as proved by the work of the Haskins group is to differentiate place of articulation. For each place of articulation a virtual 'locus' in terms of frequency values has been demonstrated with synthetic speech, notably for F_2 (e.g. Delattre *et al.*, 1955) and to a lesser extent for F_3 as well (Harris *et al.*, 1958).

As more attention has been paid to initial rather than to final transitions, the influence on the voiced-voiceless distinction of CV-transitions will be dealt with first. Subsequently, final transitions and the relation between transition and place of articulation for both types and their relative importance for Dutch will be discussed.

With the orginal synthesizer (IPOVOX I) it proved to be possible to synthesize voiceless and, within certain limits, voiced plosives without making use of transitions. Some other consonants, however, caused considerably greater difficulties (e.g. the glides /j, w/). This problem was tackled by simulating transitions through short steady-state segments: the inventory of synthetic speech sounds increased in this way, although their manipulation was rather cumbersome. For this reason a second synthesizer (IPOVOX II) was constructed with which formant transitions could be generated. In effect, the transitions followed an 'e'-power function in the frequency - time domain. The transition time was defined as the time interval necessary for the formant to move from its starting frequency (effective locus, to be called 'locus') to reach the 80% mark of the frequency range to be covered, see Fig. 6. The transition time for the two formants was not changed independently and was therefore identical throughout the tests.

Stimulated by the work of the Haskins group, this facility was used to invesigate the possible perceptual contribution of transitions towards the voiced-voiceless distinction (and place of articulation) with Dutch plosives.

A preliminary test carried out to determine the order of magnitude in which the transition had to be chosen to synthesize acceptable Dutch plosives, showed that values in excess of 30 msec, would yield glide judgments, see Fig. 7. Therefore, in all the tests involving transitions, 30 msec. constituted the upper limit of the range of values used.

CV combinations were synthesized, of which V was a two-formant vowel of the /a/ type (F_1 - 1250 c.p.s., F_2 - 1400 c.p.s.), /o/ type (F_1 - 400 c.p.s., F_2 - 800 c.p.s.) and /e/ type (F_1 - 400 c.p.s., F_2 - 3000 c.p.s.); the duration in all cases was 200 msec. whereas the pitch was made to sound natural by an F_0 inflection. These vowels were chosen so as to represent a fairly wide differentiation of formant values and to allow for a reasonable range of F_1 transitions. The consonant in the CV-syllable was synthesized as a 10 msec. noise burst with frequency peaks at either 4500 c.p.s. or 2200 c.p.s., corresponding respectively to a dental and labial place of articulation.[3]

The noise burst corresponding to a velar place of articulation was omitted on account of the absence of /k - g/ phonemic distinction in Dutch. Its inclusion would have appealed to the knowledge of a language system different from that of our Dutch listeners (cf. also fn. 1).

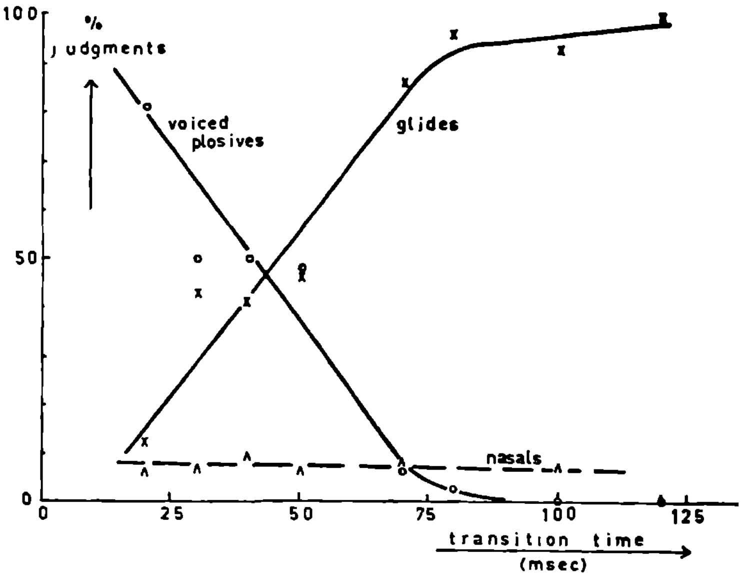


Fig. 7. Consonant judgments as a function of transition time (see Fig. 6) as obtained with synthetic CV stimuli. (Each point in the graph has been obtained from 30 judgments.)

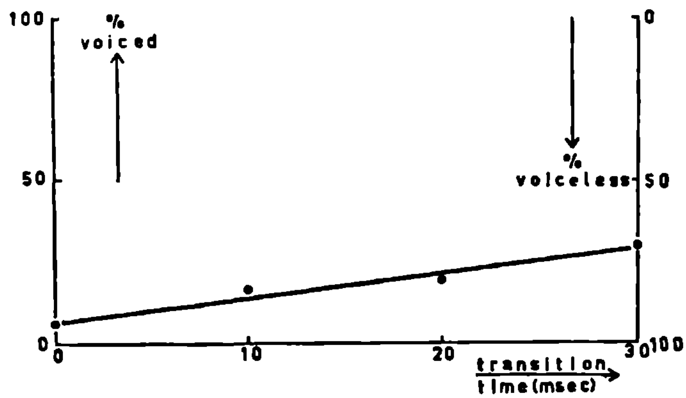
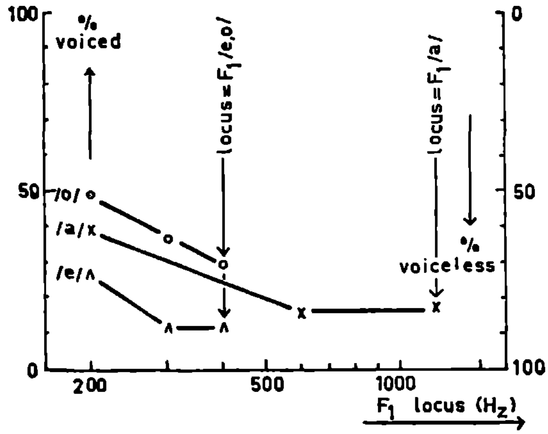


Fig. 8. Voice character as a function of transition time (see Fig. 6) as obtained with synthetic CV stimuli. (The point corresponding to transition time O has been obtained from 24 judgments and the remainder each from 272 judgments.)



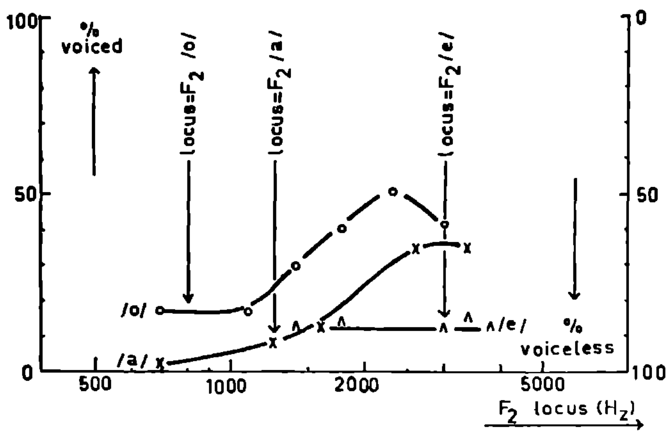


Fig. 9, 10. Voice character as a function of the locus of respectively F_1 (Fig. 9) and F_2 (Fig. 10) as obtained with synthetic CV stimuli. (Each point has been obtained from 48 judgments.)

The variables in this test were transition time, which was 10, 20 or 30 msec., locus of F_1, ranging from 200 c.p.s. to the F_1 of the following vowel in three steps, and locus of F_2, ranging from 700 c.p.s. to 3800 c.ps., depending on the vowel used maximally in eight steps and not less than five. In effect, all F_1 transitions were either zero (no transition) or negative (i.e. locus is below the formant of the vowel), the F_2 transitions being chosen so as to obtain not only zero and negative, but also positive transitions (the latter implying that the locus is higher than the vowel formant).

The results with regard to the transition time show that a small increase in the percentage of the voiced judgments occurs with longer transition times, see Fig. 8; the effect is only slight since at best it amounts to an increase of 25%.

A similar effect with respect to the influence of the frequency shift of F_1 can be observed with stimuli having a transition time of 30 msec., see Fig. 9. The biggest shift irrespective of the following vowel resulted in about 20% more voiced judgments as compared with no F_1 shift at all (zero transition). With 10 msec. and 20 msec. it was not possible to establish any influence whatever.

The extent of the positive F_2 transitions proved to have an influence on the perceived voice character of the plosives which the negative shifts did not, see Fig. 10. As no significant difference was found for the three transition times used, the judgments of the three stimulus classes are summated. The bigger positive shifts tend to cause more voiced judgments. The low F_2 position of /o/ made it feasible to present stimuli with large F_2 transitions (e.g. from 3000 c.p.s. to 800 c.p.s.), whereas the high F_2 value of /e/ did not leave room for any large shifts, but led rather to negative F_2 transitions. Consequently, the total stimulus material contained a relatively high proportion of positive F_2 transitions with respect to /o/ and a relatively low proportion with respect to /e/, with /a/ taking an intermediate position. This shows up in the high absolute value of voiced judgments in the case of /o/ in Fig. 9 compared with those for /e/ and /a/.

In order to eliminate the influence of the vowel used with respect to the F_2 transitions, the frequency values represented in Fig. 10 are normalised by dividing the locus frequency by the F_2 of the following vowel. This results in the curve given in Fig. 11, in which this quotient is plotted along the abscissa. Values < 1 represent negative transitions, 1 itself zero transition and values > 1 represent positive transitions. Fig. 11 shows more clearly than Fig. 10 that positive F_2 transitions favour voiced judgments, whereas negative ones do not.

In spite of the effect of F_2 transitions on the voiced-voiceless distinction shown with synthetic speech, they are not fully available for this purpose in actual speech, since their main task is in helping to characterise the place of articulation of the plosives. The main burden in establishing the voice character of plosives, as far as transitions are at issue, is carried by the transition time and the frequency shift of F_1.

Within the conditions of the experiment we may conclude that both transition time and frequency shift of F_1 are capable of bringing about a maximum increase or decrease of about 20% in voiced or voiceless judgments respectively. The two attributes combined, working in the same direction, may cause a shift in the judgments of 35%,

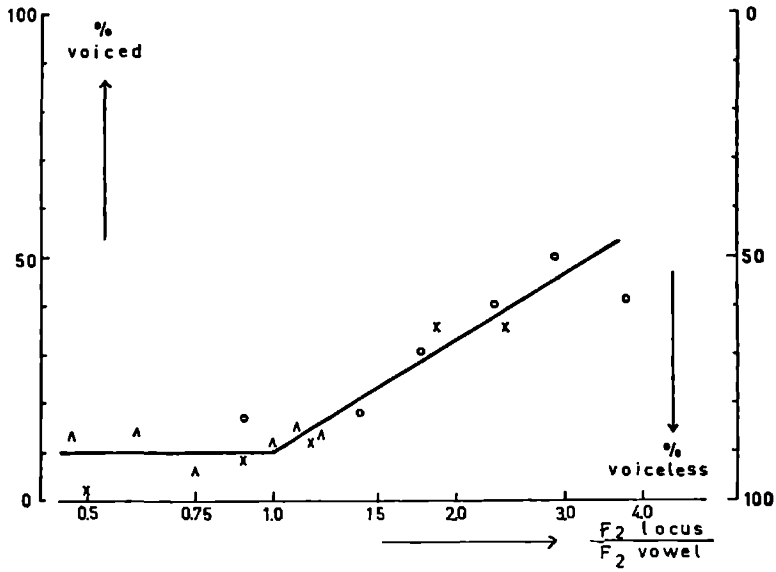


Fig. 11. Voice character as a function of the quotient obtained by dividing F_2 locus value by the steady state value of the F_2 of the following vowel. This figure is a normalization of Fig. 10.

i.e. with zero transitions the percentage of voiced judgments was found to be 4%, whereas in the optimum combined situation an average of 39% was obtained.

It thus seems legitimate to assume that these attributes have an independent influence on the voiced-voiceless distinction.

To determine the influence of final transitions, original tests were carried out of synthetic /aC/ syllables in which C again represents a labial or dental stop. The vowel /a/ was made up of a steady state segment of 160 msec. followed by a variety of formant transitions of 30 msec. duration; the consonant was synthesized by means of a noise burst of 85 msec. preceded by a silent interval of 100 msec. Listeners came up with voiceless stop responses only, which negative result might be due to the fact that normally no voiced stops occur at the end of an utterance in Dutch. Not only were the subjects unable to pronounce voiced judgments but, moreover, transitions with a large frequency shift proved to have been perceived as unnatural in this context.

To induce listeners to include voiced responses in their judgments, the stimuli were divested of the noise burst which had provided such overwhelming bias in the direction of voiceless judgments.

Under these changed experimental conditions the burstless stimuli gave rise to an increase in voiced judgments with increasing frequency shift of the F_1 transition.

In the experiments described above we found as a rule that positive F_2 transitions seem to favour /t/ or /k/ judgments, depending on the spectral composition of the noise burst, whereas negative transitions tend to cause /p/ judgments. The extent to which the transition has to be positive or negative to induce /t/ or /k/ judgments on the one side or /p/ judgments on the other, was strongly dependent on the spectral composition of the noise burst as well. No such clearly definable loci such as were found by the Haskins group could be established for Dutch.

To conclude this section on transitions, the relative contribution of initial and final transitions of the vowel will be studied. The outcome of the experiments with synthetic speech seems to indicate that initial transitions are relatively more important for the voiced-voiceless distinction than final ones. Nevertheless, since synthetic speech must be regarded as a caricature of natural speech, a control experiment with real speech was undertaken.

Series of words of the type /xəCVs/ and of the type /kVCə/ were spoken. The symbol V represents the vowels /a, o, e/, the symbol C the plosives /p, t, k, b, d/. By an electronic gating device 130 msec. of the vowel was isolated including the formant transitions to C and excluding the noise burst. The listeners were asked to guess the preceding consonant from the gated portions of the CV context and the following consonant from the gated VC context.

Of the spoken plosives in CV context, both voiced and voiceless plosives, just over 50% were correctly identified whereas again in both classes about 15% were wrongly identified with regard to the voice character only and the remainder were perceived as non-plosives, *viz.* glides, nasals, liquids or 'no consonants', see Fig 12a. Of the plosives from VC contexts both voiced and voiceless were perceived as voiceless in about 85% of the cases, see Fig. 12b. This discrepancy might be due either to the absence of a linguistic distinction between voiced and voiceless at the end of an utterance in Dutch or to lack of sufficient acoustical cues.

To determine whether this difference in outcomes of CV and VC contexts derives from an acoustical or a perceptual origin, the gated out portions were also played backwards on a tape recorder. By reversing the CV syllables one would expect that, if linguistic, i.e. distributional, rules prevail, nearly all stimuli would have to be identified with voiceless plosives as was the case wth the original VC stimuli. However, the results, see Fig. 12c, show a great number of voiced plosive judgments, which seems to rule out the preponderant effect of linguistic constraints. The fact that reversed CV stimuli do not enable listeners to distinguish original voiced from voiceless might be explained by considering that the reversed initial transitions are not identical with final transitions.

The tentative conclusion that final transitions do not contain sufficient acoustical cues for the voiced-voiceless distinction can be tested by reversing the original VC context so as to create a condition in which listeners had shown themselves to be capable of distinguishing voiced from voiceless to a reasonable extent. The results given in Fig. 12d show a slight difference only between original voiced and voiceless plosives, whereas the most characteristic feature of this figure, compared with the other three, is the

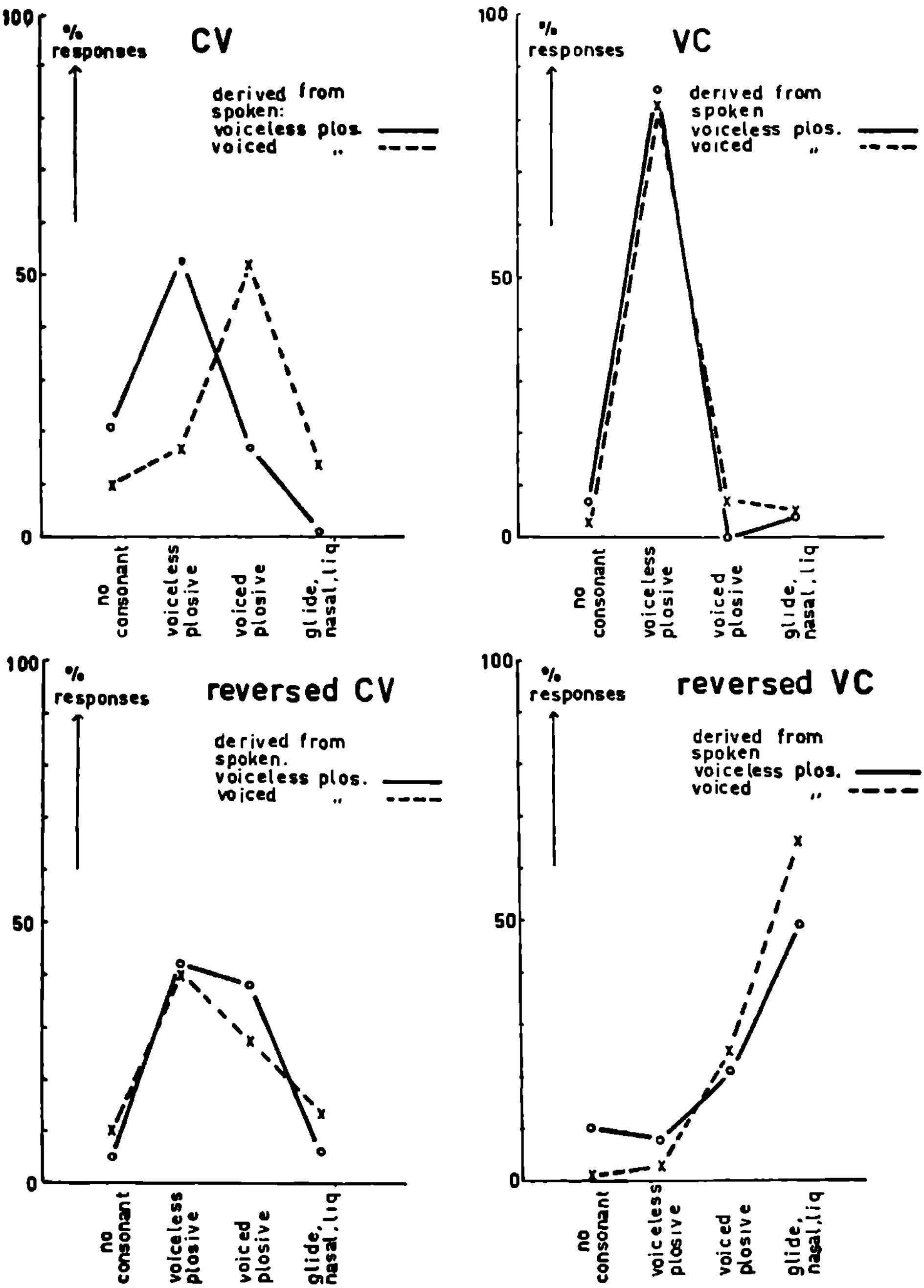


Fig. 12. Percentage of consonant judgments of gated out vowels with transitions. Fig. 12a and 12c deal with vowels obtained from words of the type /xəCVs/, Fig. 12b and 12d from words of the type /kVCə/; Fig. 12c and 12d give the responses to the reversed stimuli of Fig. 12a and 12b respectively. Each figure gives the responses to stimuli derived from 90 spoken voiceless and 60 spoken voiced plosives.

disproportionately large amount of glide judgments. This result seems to confirm the tentative conclusion about the absence of sufficient acoustical cues in final transitions.

Besides, the final transitions were experienced as longer (in time) than the initial ones, a phenomenon manifesting itself as a large number of glide judgments. As VC stimuli gave rise only to voiceless judgments, in spite of a long transition time, and the fact that the reversed CV stimuli yielded voiced responses at all, it seems justified to conclude that the frequency shift in the formants is larger with initial than with final plosives.

Note

With respect to the information contained in both initial and final transitions the frequency of F_2 of the adjoining vowel proves to have a considerable influence on the perceived place of articulation: a vowel with a high F_2 like /e/ gives rise to a predominantly labial impression, while a vowel with a low F_2 like /o/ resulted in more dental responses. A similar effect was observed with Russian subjects (Lyublinskaya, 1966, Kozhevnikov and Chistovich, 1966, p. 213).

The above tests on formant transitions concern plosives only. There are good reasons to assume that the same rules are valid for fricatives. One reason is that, if the last part of a fricative is isolated (about 10 - 30 msec.) in CV combinations together with the following vowel, a homorganic plosive with the same voice character can generally be perceived (Cohen *et al.*, 1962, Öhman, 1962). A second reason is that with synthetic speech, homorganic plosives and fricatives can be made to share the same transitions (e.g. Liberman *et al.*, 1959). While reducing the duration of fricatives by suppressing their beginning in spoken CV syllables, plosives can at times be perceived with a place of articulation different from the original one (Barrs, 1966). This might be explained by the interdependence between the starting frequency of the formant transition and the F_2 of the following vowel.

In other words " very broadly speaking . . . a fricative is a stop in which the occlusion is filled with noise " (Öhman, 1962, p. 75).

Fundamental Frequency and Intensity of Adjoining Vowels

In this section two acoustic attributes will be dealt with whose perceptual relevance has not yet been clearly established. These are fundamental frequency, F_0, of following and preceding vowels and intensity of the adjoining vowels in that order.

A difference in contour of F_0 between vowels following voiced consonants and those following voiceless ones has been noted by various investigators (Lehiste and Peterson, 1961b, Richaud and Cornut, 1962, Öhman, 1965). After voiceless consonants a gradual decrease was observed, whereas after voiced ones an initial increase was found to be followed by a decrease. The top of the F_0 contour proved to be higher after voiceless consonants (about 4.5 c.p.s. as given by House and Fairbanks, 1953; about 12 c.p.s. as can be derived from Lehiste and Peterson, 1961b).

An experiment by Fujimura (1961) with synthetic speech shows that the F_0 pattern

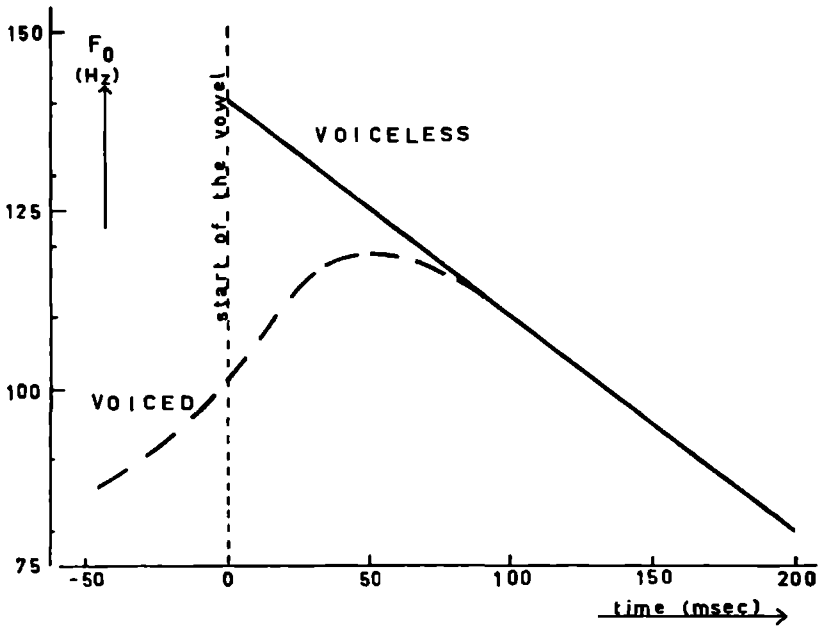


Fig. 13. Stylized representation of the influence of voice character on the F_0 contour of the following vowel. This figure is based on spectrographic data from /Ca/ syllables containing 12 voiced and 15 voiceless consonants.

has perceptual consequences. The presence of an F_0 pattern (a rise from 70 to 100 c.p.s.) proved to be favourable to the perception of voiced plosives.

Acoustic measurements for Dutch show comparable differences in F_0 contours, see Fig. 13. The top frequency after voiceless consonants was found to be 6 c.p.s. (s = 2 c.p.s.) higher than that of F_0 after voiced ones.

The influence of the voice character of consonants on the F_0 of preceding vowels is less definite. Measurements by Lehiste and Peterson (1961b) show no significant difference. However, a calculation based on the data given by them shows that the mean F_0 value of vowels preceding voiced consonants is 2 c.p.s. higher than that of vowels preceding voiceless consonants. Similar measurements for Dutch yield a corresponding difference of 2.5 c.p.s. (s = 1.5 c.p.s.) in favour of the voiced situation.

A higher intensity in vowels adjoining voiced consonants than in vowels adjoining voiceless consonants was measured by Lehiste and Peterson (1959). House and Fairbanks (1953) observed the same phenomenon with vowels embedded in voiced contexts as compared with those in voiceless surroundings.

For Dutch the amplitudes of vowels preceding and following voiced and voiceless consonants, both spoken and whispered, were measured from oscillographic recordings.

The results are expressed in terms of the mean ratio obtained by dividing the amplitudes of vowels in the voiced situation by that of vowels in the voiceless one in Table 1.

TABLE 1

Amplitude ratio for vowels adjoining voiced versus voiceless consonants

	NORMAL SPEECH	WHISPER
PRECEDING VOWEL	1.18	1.04
FOLLOWING VOWEL	0.97	1.21

(s = 0.06)

These results and those obtained by others seem to show a tendency for the amplitude of vowels adjoining voiced consonants to be slightly higher than that of vowels in voiceless situations.

REFERENCES

ABRAMSON, A. S. and LISKER, L. (1965). Voice onset time in stop consonants: acoustic analysis and synthesis. Status Rep. Sp. Res., Haskins Labs. (New York), No. 3, 1.1.

BARRS, J. T. (1966). Identification of voiceless initial fricatives versus modification of their duration. *Linguistics*, 21, 5.

BELASCO, S. (1953). The influence of force of articulation of consonants on vowel duration. *J. acoust. Soc. Amer.*, **25**, 1015.

COHEN, A., SCHOUTEN, J. F. and 'T HART, J. (1962). Contribution of the time parameter to the perception of speech. *Proc. 4th Int. Congr. Phon. Sci., Helsinki* (The Hague), 555.

COHEN, A. and 'T HART, J. (1962a). Speech synthesis of steady state segments. *Sp. Commun. Sem.* (Stockholm), F 1.

COHEN, A. and 'T HART, J. (1962b). Segmentation of the speech continuum. *Proc. 4th Int. Congr. Acoust.* (Copenhagen), G 5.

COHEN, A., SLIS, I. H. and 'T HART, J. (1963). Perceptual tolerances of isolated Dutch vowels. *Phonetica*, **9**, 65.

COHEN, A. (1964). Phonetic research. *Philips Techn. Rev.*, **25**, 43.

COHEN, A., SLIS, I. H. and 'T HART, J. (1967). On tolerance and intolerance in vowel perception. *Phonetica*, **16**, 65.

COOPER, F. S., DELATTRE, P. C., LIBERMAN, A. M. and BORST, J. M. (1952). Some experiments on the perception of synthetic sounds. *J. acoust. Soc. Amer.*, **24**, 597.

DELATTRE, P. C., LIBERMAN, A. M. and COOPER, F. S. (1955). Acoustic loci and transitional cues for consonants. *J. acoust. Soc. Amer.*, **27**, 769.

DELATTRE, P. C. (1962a). Le jeu des transitions de formants et la perception des consonnes. *Proc. 4th Int. Congr. Phon. Sci., Helsinki* (The Hague), 407.

DELATTRE, P. C. (1962b). Some factors of vowel duration and their cross-linguistic validity. *J. acoust. Soc. Amer.*, 34, 1141.

DENES, P. (1955). Effect of duration on the perception of voicing. *J. acoust. Soc. Amer.*, 27, 761.

FISCHER-JØRGENSEN, E. (1954). Acoustic analysis of stop consonants. In *Miscellanea Phonetica II* (London), 42.

FISCHER-JØRGENSEN, E. (1964). Sound duration and place of articulation. *Zeitschr. f. Phon.*, 17, 175.

FORREZ, G. (1966). Relevante parameters van de stemhebbende fricatief /z/. *Inst. Perceptie Onderzoek Verslag* (Eindhoven), No. 55.

FUJIMURA, O. (1961). Some synthesis experiments on stop consonants in initial position. M.I.T. Quart. Prog. Rep., No. 61, 153.

HALLE, M., HUGHES, G. W. and RADLEY, J. P. A. (1957). Acoustic properties of stop consonants. *J. acoust. Soc. Amer.*, 29, 107.

HARRIS, K. S., HOFFMANN, H. S., LIBERMAN, A. M., DELATTRE, P. C. and COOPER, F. S. (1958). Effect of third formant transitions on the perception of voiced stop consonants. *J. acoust. Soc. Amer.*, 30, 122.

'T HART, J. and COHEN, A. (1964). Gating techniques as an aid in speech analysis. *Language and Speech*, 7, 22.

HOUSE, A. S. and FAIRBANKS, G. (1953). The influence of consonant environment upon the secondary acoustical characteristics of vowels. *J. acoust. Soc. Amer.*, 25, 105.

JASSEM, W. (1962). The acoustics of consonants. *Proc. 4th Int. Congr. Phon. Sci.* (The Hague), 50.

KOZHEVNIKOV, V. A. and CHISTOVICH, L. (1966). *Speech: Articulation and Perception.* Translation of *Rech: Artikulyatsiya i Vospriyatiye* (Moscow-Leningrad), 1965.

LEHISTE, I. and PETERSON, G. (1959). Vowel amplitude and phonemic stress in American English. *J. acoust. Soc. Amer.*, 31, 428.

LEHISTE, I. and PETERSON, G. (1961a). Transitions, glides and diphthongs. *J. acoust. Soc. Amer.*, 33, 268.

LEHISTE, I. and PETERSON, G. (1961b). Some basic considerations in the analysis of intonation. *J. acoust. Soc. Amer.*, 33, 419.

LIBERMAN, A. M., DELATTRE, P. C., GERSTMAN, L. J. and COOPER, F. S. (1956). Tempo of frequency change as a cue for distinguishing classes of speech sounds. *J. exp. Psychol.*, 52, 127.

LIBERMAN, A. M., DELATTRE, P. C. and COOPER, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, 1, 153.

LIBERMAN, A. M., INGEMANN, F., LISKER, L., DELATTRE, P. C. and COOPER, F. S. (1959). Minimal rules for synthesizing speech. *J. acoust. Soc. Amer.*, 31, 1490.

LIBERMAN, A. M., HARRIS, K. S., EIMAS, P., LISKER, L. and BASTIAN, J. (1961). An effect of learning on speech perception: the discrimination of durations of silence with and without phonemic significance. *Language and Speech*, 4, 175.

LISKER, L. and ABRAMSON, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384.

LISKER, L. and ABRAMSON, A. S. (1965). Voice onset time in the production and perception of English stops. Status Rep. Sp. Res., Haskins Labs. (New York), No. 1, 3.1.

LYUBLINSKAYA, V. V. (1966). Recognition of articulation cues in stop consonants in transition from vowel to consonant. *Soviet Physics—Acoustics*, 12, 185.

ÖHMAN, S. E. G. (1962). On the perception of Swedish consonants in intervocalic position. Rep. No. 25, Royal Inst. Techn. (Stockholm).

ÖHMAN, S. E. G. (1965). On the co-ordination of articulatory and phonatory activity in the production of Swedish tonal accents. STL-QPSR, Royal Inst. Techn. (Stockholm), No. 2, 14.

PETERSON, G. and LEHISTE, I. (1960). Duration of syllable nuclei in English. *J. acoust. Soc. Amer.*, 32, 693.

RICHAUD, M. C. and CORNUT, G. (1962). Variations du fonctionnement laryngé en fonction des phonèmes et des transitions phonétiques dans la voix parlée et chantée. *Proc. 12th Int. Congr. Speech & Voice Ther.* (Padua), 166.

SHARF, D. J. (1964). Vowel duration in whispered and in normal speech. *Language and Speech*, 7, 89.

SLIS, I. H. (1966). A model for the distinction between voiceless and voiced consonants. Annual Progr. Rep., Inst. Percept. Res. (Eindhoven), No. 1, 40.

SLIS, I. H. (1967). What causes the voiced-voiceless distinction ? Annual Progr. Rep., Inst. Percept. Res. (Eindhoven), No. 2, 71.

STREVENS, P. (1960). Spectra of fricative noise in human speech. *Language and Speech*, 3, 32.

TSCHESCHNER, W. (1966). Untersuchungsergebnisse der Sprachanalyse. *Zeitschr. f. Phon.*, 19, 141.

VIERREGGE, W. H. (1966). Die akustische Struktur der Plosivlaute. *Forschungsbericht*, Inst. Phon. (Bonn), 66/3, 48.

WILLEMS, L. F. (1966). IPOVOX II: a speech synthesizer. Annual Progr. Rep., Inst. Percept. Res. (Eindhoven), No. 1, 120.

ZIMMERMANN, S. A. and SAPON, S. M. (1958). Note on vowel duration seen cross-linguistically. *J. acoust. Soc. Amer.*, 30, 152.

CHAPTER 3

ON THE COMPLEX REGULATING THE VOICED-VOICELESS DISTINCTION II

I. H. Slis and A. Cohen

Reprinted from:

LANGUAGE AND SPEECH
Vol. 12, part 3, 1969, 137-155

ON THE COMPLEX REGULATING THE VOICED-VOICELESS DISTINCTION II*

I. H. SLIS and A. COHEN
Institute for Perception Research, Eindhoven

TIME STRUCTURE OF THE ACOUSTIC ATTRIBUTES

An attempt will be made to establish the inter-relation in time of the various acoustic attributes as described in the previous section. This will be done first in CV contexts and subsequently extended to include a preceding vowel, thus constituting a VCV context.

In the second part of this section the relative perceptual contribution of the attributes will be discussed with the help of a variety of experiments including whispered speech.

The data in this section will apply in principle both to plosives and fricatives. Where no actual data on the fricatives are available an account will be given of the repercussions of the results obtained in studying plosives.

Thus far, each attribute has been investigated singly with respect to its contribution towards the voiced-voiceless distinction. Varying the parameter values of a single acoustic attribute of synthetic speech would in most cases make a modest contribution towards the question of voicing. Therefore it seems justified to assume that in normal speech these various factors do not operate independently of one another.

A stylised synthesis of the attributes in time

In studying CV contexts the following successive acoustical events can be observed where the C's stand for a voiced plosive (/b, d/): (1) a voice lead (also known as 'Blählaut' and visible on a spectrogram as a voice bar preceding the noise burst) of about 80 msec. (2) a short noise burst of about 5 msec. (3) vowel with extensive initial formant transitions of F_1 and F_2 with an effective duration of about 30 msec. With voiceless plosives a similar description would give the following succession: (1) a long noise burst of about 20 msec. (2) formant transitions over a small frequency range of about 10 msec.

A comparison of these two series of events as shown in Fig. 14 shows that, if the moment of the beginning of the noise burst is defined as $t = 0$ msec., the voice lead in (a) (voiced) covering the interval from -80 msec. to 0 msec. is absent in (b) (voiceless). Next, the part between 0 msec. and 20 msec. is filled with a noise burst lasting from 0 to 5 msec. and part of the formant transitions leading into the vowel in (a), whereas it is wholly taken up by a noise burst in (b). In fact, this difference boils down

* *The first part of this paper appeared in* Language and Speech, *1969, 12, 80.*

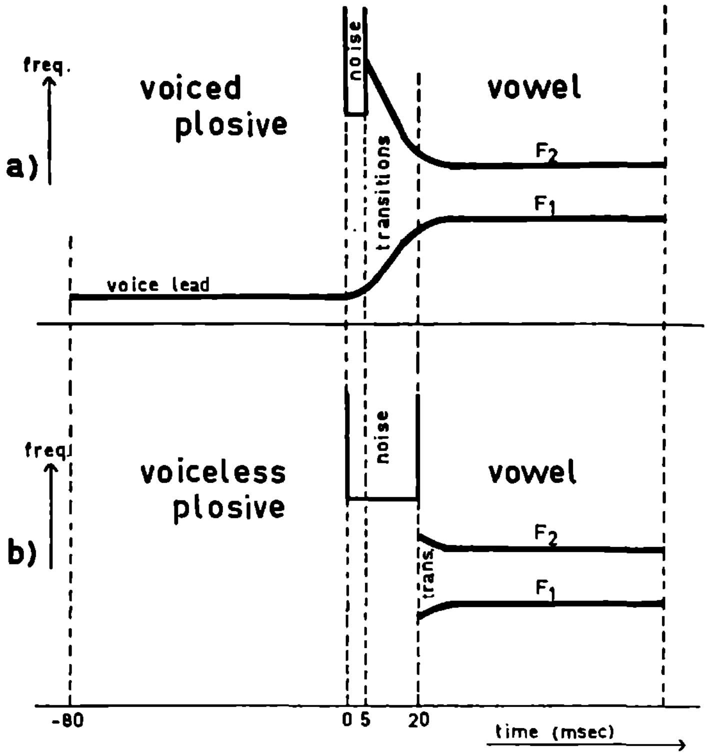


Fig. 14. Stylised representation of CV syllables.
(a) with a voiced plosive, containing voice lead, noise burst, transitions and vowel;
(b) with a voiceless plosive, containing noise burst, transitions and vowel.

to a delay in the moment of voice onset in (b) compared with (a). As a consequence, the part of the formant transitions before $t = 20$ msec. in (a) fails to turn up in (b) as vowel formant transitions; the realized part of the transitions in voiceless plosives, therefore shorter (in time), shows a less extensive frequency shift than that of the voiced plosives.

In this stylised representation no account is given of the possible presence of transition-like features in the noise part as may be seen in spectrographic recordings (see Fant *et al.* 1963); moreover, use has been made of this spectrographic finding by filtering noise through vowel formant filters in synthesizing plosives (e.g. Fujimura, 1961a; Liberman *et al.*, 1958). The reason why such noise transitions are not included in the stylised version, represented here, is that we have not been able to find sufficient evidence in our recordings for Dutch.

TABLE 2

	MEAN DURATION OF PRECEDING VOWEL a	MEAN DURATION OF SILENT INTERVAL b	a + b
/bə($^{b}_{d}$)œt/	131 msec.	134 msec.	265 msec.
/bə($^{p}_{t}$)œt/	106	168	274
/ba($^{b}_{d}$)ə/	250	85	335
/ba($^{p}_{t}$)ə/	220	112	332
/bɑ($^{b}_{d}$)ə/	162	105	267
/bɑ($^{p}_{t}$)ə/	129	140	269

As a result of the delay in voice onset in (b) (voiceless), the following vowel is shorter than that in (a) (voiced). That this is not a mere artefact showing how the difference between voiced and voiceless plosives is illustrated can be substantiated by evidence from literature (Fischer-Jørgensen, 1964).

Apart from the possible compensation in terms of vowel length with regard to the vowel following the plosive, such a compensation can be shown to occur in Dutch with respect to the preceding vowel. When the mean differences in duration of the preceding vowels and that of the silent intervals are compared they seem to balance one another; this means that adding the duration of vowels and silent intervals of voiced and voiceless plosives in similar contexts would give equal values. That this is, in fact, the case can be seen from Table 2 in which the average times are given for whispering and normal speaking (two speakers).

It is possible to extend the stylised representation of Fig. 14 with a preceding vowel as shown in Fig. 15: the preceding vowel ends at $t=0$ — s.i. in (a) (voiced) and at $t = (0$ — s.i.) — 30 msec. in (b) (voiceless). In (a) the voice bar can be considered as a continuation of the first formant during the silent interval. That this interval is filled with a voice bar has been observed for most languages (Lisker and Abramson, 1964). As the sound level of this voice bar is low compared with the adjoining vowels it seems justified to maintain the term "silent" interval.

Comparison of (a) and (b) shows that in (b) the voice stops earlier than the silent interval of (a) starts: this implies that the formant transitions do not develop fully in (b) whereas they do in (a) (Fig. 15).

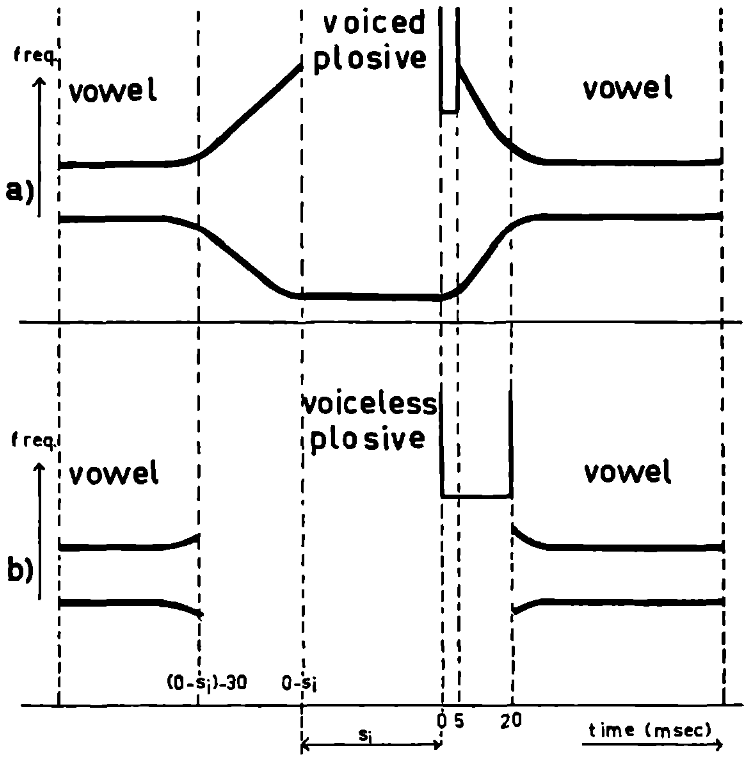


Fig. 15. Stylised representation of VCV words
(a) with a voiced plosive;
(b) with a voicessless plosive.

Looking at the overall picture, the most characteristic difference between voiced and and voiceless is the continuation of voice activity in (a) and its early cessation and delayed start in (b), (cf. also Öhman, 1965, Fig. II B2).

A similar descriptive device can be used to compare the time relations of VCV structures containing fricatives, resulting in Fig. 16. Here again in the lower part of the figure a voiceless interval can be assumed whereas a continuation of voice occurs in the upper part of the figure; and again, as with plosives, the voice stops early and starts late again in the voiceless situation as compared with the beginning and end of the fricative interval of the voiced situation. Data obtained with measurements on fricatives, as mentioned in the previous section, indicate a mean difference of 40 msec. in the preceding vowels and of 50 msec. in the fricative intervals. The

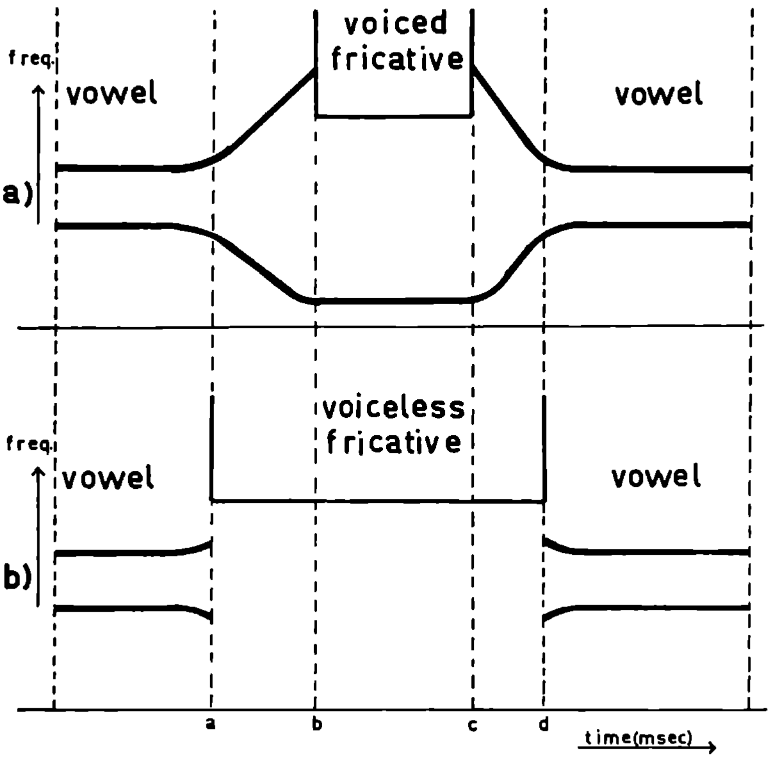


Fig. 16. Stylised representation of VCV words
(a) with a voiced fricative
(b) with a voiceless fricative
a, b, c, d on the base line indicate the moment the consonant starts (a, b) or stops (c, d) with voiced (b, c) or voiceless (a, d) fricatives.

slight effect of over-compensation (about 10 msec.) seems to be counterbalanced by a reduction of the second vowel.

The hypothesized compensation is confirmed from measurements of the total duration of pairs of words differing only in voice character of the consonants at issue. The mean of the differences in duration of the words constituting a pair was found to be 0.5 msec. ($s = 4$ msec.), which is negligibly small.

Perceptual relations of the attributes

Having first given an inventory of the attributes and their relation in time, we will now explore the voiced-voiceless distinction by selecting sub-sets of attributes to establish their relative perceptual influence. To this end a number of heterogeneous

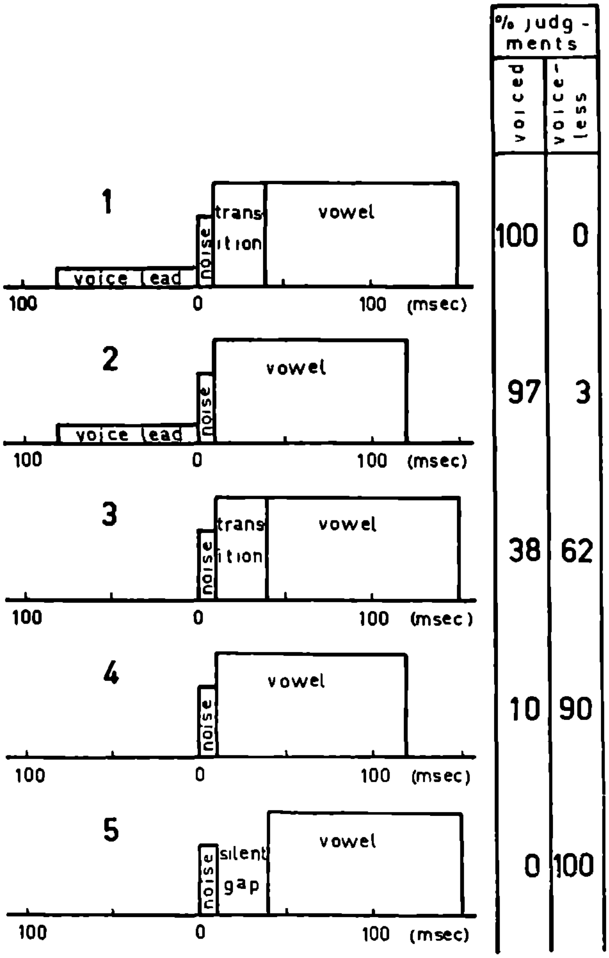


Fig. 17. Survey of synthetic stimuli used in a test to compare the contribution of the various acoustic attributes with respect to the voiced-voiceless distinction. To the right the results of a listening test are given in terms of percentage of voiced and voiceless judgements for each type of stimulus. (With each stimulus type 18 judgments were obtained.)

TABLE 3

Percentage of correction identification of isolated consonants and vowels with consonant transition.

	CONSONANT	VOWEL WITH TRANSITION
Voiceless fricatives /s, f, x/	91%	24%
Voiced fricatives /z, v/	84	51
Voiceless plosives /p, t, k/	79	22
Voiced plosives /b, d/	62	25
Nasals, liquids /m, n, l, r/	69	40
Glides /w, j/	69	76

experiments will be described, which originally were carried out independently of one another, including some by other investigators. Thus we hope to obtain a better understanding of the nature of the mechanism underlying the voiced-voiceless distinction.

A test was arranged to check the influence of the exchange of some of the leading attributes determining the voice character of plosives with synthetic speech. Five types of stimuli were generated of CV structure in which the noise burst and steady state of the vowel were made so as to correspond respectively to labial or dental plosives and vowels /a, i, u/. Four of these stimuli are the possible combinations obtained by adding or omitting voice lead and transitions (synthesized by means of short steady state segments); a fifth stimulus was added in which a silent gap took the place of the transition of stimulus 3, see Fig. 17. The results based on the judgments of three subjects are also given in this figure (right).

The voice lead seems to be a stronger cue than the transitions in Dutch under these experimental conditions. The gap introduced in stimulus 5, delaying the voice onset, resulted in 100% voiceless change-over as compared with stimulus 4.[1]

A test with spoken CV syllables was carried out to establish the relative perceptual weight of the consonant *v.* the vowel part, C standing for all initial consonants except /h/ and V for /a, u, i, ø/. Either the consonant or the vowel, including transitions, could be presented as a stimulus to four subjects. The subjects were supposed to identify the consonant; where C was actually present in the stimulus the results were scored as before, whereas in the other type of stimulus the correct place of articulation was given one point and if the correct voice character was also perceived two marks were given. The results are given in Table 3.

[1] *This experiment shows a strong resemblance to one reported on p. 84 in which 10 msec. of the friction noise together with the following vowels were isolated from spoken CV syllables to which a voice lead could be added; the stimuli derived from the originally spoken voiced consonants are comparable to stimulus 1 and 3 and those derived from their voiceless counterparts are comparable to stimulus 2 and 4 in Fig. 17.*

TABLE 4

Degree of annoyance caused by the introduction of a 30 msec. gap at the beginning of the vowel transition as evaluated by 4 subjects.

	ISOLATED CV SYLLABLES	WORDS FROM RUNNING SPEECH
Control experiment without gap	0.3	0.2
Voiceless fricatives (/s, f, x/)	1.4	2.0
Voiced fricatives (/z, v/)	2.4	2.2
Nasals, liquids (/n, m, l, r/)	2.7	2.8
Glides (/j, w/)	3.2	3.1
Control experiment with gap in vowel	3.0	3.7

It seems that the vowel transitions contain more information with voiced consonants than they do with voiceless ones, and that the friction part of voiceless consonants contains more information than the friction part of voiced ones. With fricatives more information appears to be present in the friction than is the case with plosives.

To determine the relative contribution of the part where consonant and vowel meet in CV syllables, 50 msec. (20 msec. of C, 30 msec. of V) of the stimuli described above were presented first. No appreciable difference was found to exist between voiced and voiceless consonants, since a very high identification score was obtained in all cases (around 90%). As a next step the first 30 msec. of the vowel part was suppressed, resulting in an interruption of the recorded speech signal, consisting of CV syllables spoken in isolation and comparable to those described in the preceding section, with the exception of the plosives. In another situation, obtained from running speech, giving stimuli in the shape of complete words, a gap was introduced in —CV— contexts in which C was not necessarily initial. As identification was nearly perfect in both situations the subjects were instructed to express the nuisance value of the gaps on a five-point scale, ranging from 0 ('no gap') to 4 ('a very annoying gap'). The results are summarised in Table 4. This table also gives the results of a control experiment in which there was no gap and in which there was a gap in the middle of the vowel.

The nuisance value seems to be inversely proportional to the acoustic contrast between consonant and vowel. In effect, a suppression of the transition with voiceless fricatives proved to be hard to detect. This finding correlates well with the acoustic phenomenon that a gap of 10-20 msec. can often be observed in oscillographic recordings of voiceless fricatives in CV contexts.

Since all the stimuli could easily be identified, and the gaps with fricatives were hardly noticed or perceived as small, it seems legitimate to conclude that the relative contribution of the transitions with fricatives to the voiced-voiceless distinction is negligible.

TABLE 5

	PERCENTAGE ERRORS IN VOICE CHARACTER	
	CV	VCV
Voiceless fricatives (/s, f/)	22	12
Voiceless plosives (/t, p/)	18	4
Voiced fricatives (/z, v/)	36	35
Voiced plosives (/d, b/)	48	57

Commutation tests involving Danish plosives (Fischer-Jørgensen, 1956) show that the identification of voiceless /p, t, k/ is hardly affected by the following vowel and is almost carried by the noise burst proper. However, with voiced /b, d, ɡ/ the vowel plays a predominant part in consonant identification.

One conclusion from the work of Fischer-Jørgensen could be that the main information for identification of voiceless plosives resides in the noise burst, whereas with voiced plosives it is largely contained in the vowel (transition) in Danish.

A similar test done on Swedish fricatives (Martony, 1962) shows that " the friction is the most important cue for all fricative sounds " (p.5).

These results seem to confirm those obtained with synthetic speech through which it is possible to generate good voiceless plosives without vowel transitions, unlike voiced ones where transitions can hardly be done without if acceptable quality is required. What has been said of voiceless plosives applies *a fortiori* to the voiceless fricatives. As for the voiced fricatives, although the transitions play a much less important part than they do with the voiced plosives, they have " . . . . some cue value . . . . " (Delattre *et al.*, 1962, p. 113).

Studying the relative contribution of the voice bar and the final part of the preceding vowel with voiced plosives in VCV contexts, Öhman (1962 a, b) showed that removal of either cue in itself did not lead to a change in voice character, whereas a removal of both cues did.

With a preliminary experiment on spoken Dutch embedded plosives we observed that, by adding or removing a voice bar only, a change of voice character could be brought about in most cases.

Measurements of the time cues controlling the voiced-voiceless distinction in whispered speech were not significantly different from those carried out on normal speech (cf. section I.). Consequently it seems justified to consider whispered speech as normal speech minus voice, leaving the time structure and probably transitional cues intact. In this way it becomes possible to investigate the relative perceptual contribution of the time cues to the voiced-voiceless distinction.

A perceptual test was done with whispered CV and VCV combinations, where C stands for /p, t, f, s, b, d, v, z/ and V for /a, ɑ, o, u, e, i/. Three speakers and four listeners participated in this test.

The results summarised in Table 5 show that the voice character is less distinct with whispered speech than with normal speech. (In the latter, identification was 100%.) The fact that the majority of errors were made with the voiced consonants indicates that the absence of voice is highly detrimental to the voiced-voiceless distinction. This outcome might be considered as additional evidence that the vowel part carries the major share of the information in the case of voiced consonants. Since with voiceless consonants the major burden is carried by the friction noise, which is not affected by whispering, it should cause no surprise that they are relatively well identified.

Some evidence of the effect of time and intensity cues is apparent from a comparison of voiced and voiceless consonants both in CV and in VCV contexts. The difference between voiceless consonants in CV and VCV might be due to time cues inherent in the preceding vowel.

In this section we believe we have established the interconnection of time cues controlling the voiced-voiceless distinction and proved that this distinction is not confined to a segment corresponding to a single consonant phoneme but rather exerting its influence over immediately adjoining segments at least. Moreover, the presence or absence of voice can not only be regarded as the major point of departure in drawing up a stylised model based on acoustic attributes but has also proved to be a major cue in perception.

Since " voice " obtrudes itself as a factor regulating the complex of acoustical and perceptual attributes, we shall try to explain this phenomenon by considering the articulatory mechanism on which the presence or absence of voice is based.

Model of articulation

In this section we shall explore the articulatory mechanism which may be held responsible for the voice character of the consonants under discussion. In an attempt to reduce the complex to a single cause, a number of aspects will be discussed, such as the relation between voice and pressure drop across the glottis and the build-up of the intra-oral pressure.

In the concluding sections the model will be developed and its consequences illustrated.

Voice-pressure relation

For various reasons it is possible to conclude that the majority of the symptoms affecting the voiced-voiceless difference are reducible to an advanced voice stop and a delayed voice onset with voiceless consonants; with voiced consonants often no voice stop occurs at all.

The assumption that the action of the vocal cords is stopped and started directly by a nerve command is not feasible, since the time necessary to separate the vocal cords (80-90 msec.) and to bring them together again (75-90 msec.) is shown to be in the order of 155- 180 msec. (Ventzov, 1966). This time is longer than the voiceless interval.

A more likely explanation is the one based on the myoelastic-aerodynamic theory of voice production (van den Berg, 1953, 1957, 1958): a pressure drop across the glottis gives rise to an airstream through the glottis; therefore the pressure in the glottis decreases, the vocal cords are sucked together and the glottis closes; due to the pressure drop across the glottis the vocal cords are pressed apart and the glottis opens again, etc. . . . The necessity of this pressure drop has been demonstrated with patients (Gougerot *et al.*, 1958) and tracheotomised dogs (Vallancien, 1958; Hast, 1962).

As higher intra-oral pressures are measured with voiceless consonants than with voiced ones, it is possible for the pressure drop across the glottis with voiceless consonants to be too small to support vibration (e.g. Damsté, 1961; Fischer-Jørgensen, 1963; Malécot, 1966; Subtelny *et al.*, 1966; Slis and Damsté, 1967). If the pressure drop is responsible for the interruption of the action of the vocal cords, it must be possible to bring them into action again by increasing the pressure drop, for instance by decreasing the intra-oral pressure. This has been effectuated by Ventzov (1966). The results show a considerably shorter voiceless interval or no voiceless interval at all with a decreased intra-oral pressure. In a similar experiment Kozhevnikov and Chistovich (1966) show that if a subject speaks with an open tube in his mouth, the moment of voice onset with voiceless plosives is advanced. On this basis they assume that the vocal cords are already in a phonation position before the moment of voice onset with voiceless plosives.

Pressure measurements reported by Lisker (1965) show less significant results, although the voiceless plosives tend to higher intra-oral pressures. An explanation given by Lisker is that the pressure phenomena are a consequence of mode of larynx operation (open or closed glottis) and not the other way round. For the time being we shall follow the first line of thought, namely, the cessation of vocal action is a consequence of a decreased pressure drop.

Intra-oral pressure

The way the intra-oral pressure is built up differs for voiced and voiceless consonants and depends on context. With embedded voiceless plosives the pressure rises quickly and remains approximately constant during the remainder of the closed interval; with voiced plosives a gradual rise occurs on which oscillations caused by the action of the vocal cords are superposed. This rise extends over the whole closed period (Fischer-Jørgensen, 1963; Stetson, 1951; Lisker, 1965). (With initial plosives no differences in the mode of the pressure build-up are registered between voiced and voiceless.)

Several possible causes have been suggested to explain this difference in behaviour of pressure build-up and in peak value of the pressure.

Lisker (1965) states that: ". . . we are entitled to suppose that a difference in supraglottal pressure is a consequence of a difference in mode of laryngeal operation, rather than the sign of an independent fortis-lenis contrast ". In another article in the same Haskins report an indication can be found of what Lisker means by " mode of laryngeal operation " (Abramson, Lisker and Cooper, 1965).

Preliminary tests with transillumination of the glottis show differences in laryngeal aperture; voiceless plosives show a wide glottal opening, whereas with voiced plosives the vocal cords remain in phonatory position. In a similar way Malécot and Peebles (1965) show that a relatively higher opening of the glottis occurs with voiceless plosives than with voiced ones; this holds for speaking as well as whispering (see also Slis and Damsté, 1967).

In studying the glottal behaviour with an endoscope, Borel-Maisonny (1958) established that during voiceless consonants in French the glottis is not necessarily closed.

During Dutch voiceless consonants the glottis is a " narrow fusiform slit " as observed by means of a phonolaryngoscope by Eijkman (1933, p. 219). One may safely conclude that there is a difference in the behaviour of the vocal cords and therefore an influence of this behaviour on the intra-oral pressure has to be assumed.

Throughout the literature references may be found to a correlation between height of the larynx and voice character of consonants to the effect that ' high ' correlates with voiceless and ' low ' with voiced (e.g. Sievers, 1881; Roudet, 1910; Grammont, 1939; Fischer-Jørgensen, 1963). According to some investigators this difference in height is due to a downward movement for voiced (4 - 5 mm., cf. Hudgins and Stetson, 1935); other investigators point to an upward movement during voiceless consonants (Perkell, 1965a).

On the basis of this phenomenon it has to be assumed that the pharynx is relatively large for voiced consonants as compared with voiceless ones. It stands to reason that this difference in pharynx volume has consequences for the build-up of the intral-oral pressure.

Another mechanism that affects the volume of the pharynx is the musculature in the pharynx wall. Hudgins and Stetson (1935) write that ". . . the entire musculature of the pharynx region . . . is tense during the surd occlusion ". This tension can be regarded as responsible for the fact that the increase in the pharynx width is less than with voiced, as observed by Perkell (1965a, b) through X-ray films; the extra increase in volume with voiced plosives seems approximately equal to the volume of air necessary to permit vocal vibration.

The contraction of this musculature is made use of in the instruction of oesophageal speech of laryngectomized patients (Damsté, 1958). This contraction functions as a pump-like mechanism; " it produces a counter pressure that stops the vibrations of the glottis while at the same time it provides the explosive force for /p,t,k/" (Damsté, 1959).

Drawing up the model

It is possible to combine these physiological events into one model by assuming that the pharyngeal musculature provides the origin of the observed phenomena.

In the voiceless situation a tightening of the pharyngeal musculature leads to a contraction of the pharynx wall and a rising of the larynx. Consequently the volume of the pharynx decreases considerably because of a narrowing of its width and a raising

TABLE 6

PROPERTY DIFFERING IN THE WORD PAIRS	MEAN DIFFERENCE IN DURATION OF THE WORDS	S.D. OF THE MEAN
Voice character (voiceless > voiced)	0.5 msec.	4 msec.
Manner of articulation (fricative > plosive)	23.3	5.5
Place of articulation (dental > labial)	10.5	3.5

of its bottom. As part of this musculature has an influence on the inner state of the larynx (Zenker, 1966) it is probable that it has an influence on the vocal cords as well. Therefore the difference observed in the glottal area with voiced and voiceless consonants can also be referred back to the pharyngeal musculature.

The small volume of the pharyngeal-oral cavity combined with a low resistance in the glottis, because of the more or less abducted condition of the vocal cords results in a rapid build-up of the intra-oral pressure with voiceless consonants; the large elastic (not tense) pharyngeal-oral cavity combined with a high resistance in the phonating glottis results in a gradual build-up of the pressure with the voiced consonants.

Measurements of total duration of words of the types /bə-C-œt/, /ba-C-ə/ and /bɑ-C-ə/ show that, with word-pairs differing in voice character only, no significant difference in mean duration is observed. If pairs are compared differing in mode of articulation, words containing fricatives prove to be significantly longer than those with plosives, and words containing dentals longer than those with labials (see Table 6; cf. Kozhevnikov and Chistovich, 1966, p. 107).

These results show that if there is a change in the mode of operation of muscle groups, responsible for the closing and opening movements of the mouth (fricative *v.* plosive and dental *v.* labial), its consequences manifest themselves in a difference of duration. A difference in voice character alone does not seem to reveal such a difference in duration; therefore it seems justified to assume that the voiced-voiceless distinction is controlled by a muscle or muscles operating quite independently of the closing musculature.

The constrictor muscle constitutes a likely choice since it does not interfere with either place or manner of articulation and exerts influence on the position and the condition of the larynx as shown by Zenker (1966).

The implications of the model

A survey of the model and its implications is given in Fig. 18. The top part of the diagram represents the physiological events, *viz.* the immediate influence of the pharyngeal constrictor muscle on the pharyngeal wall and larynx position and their

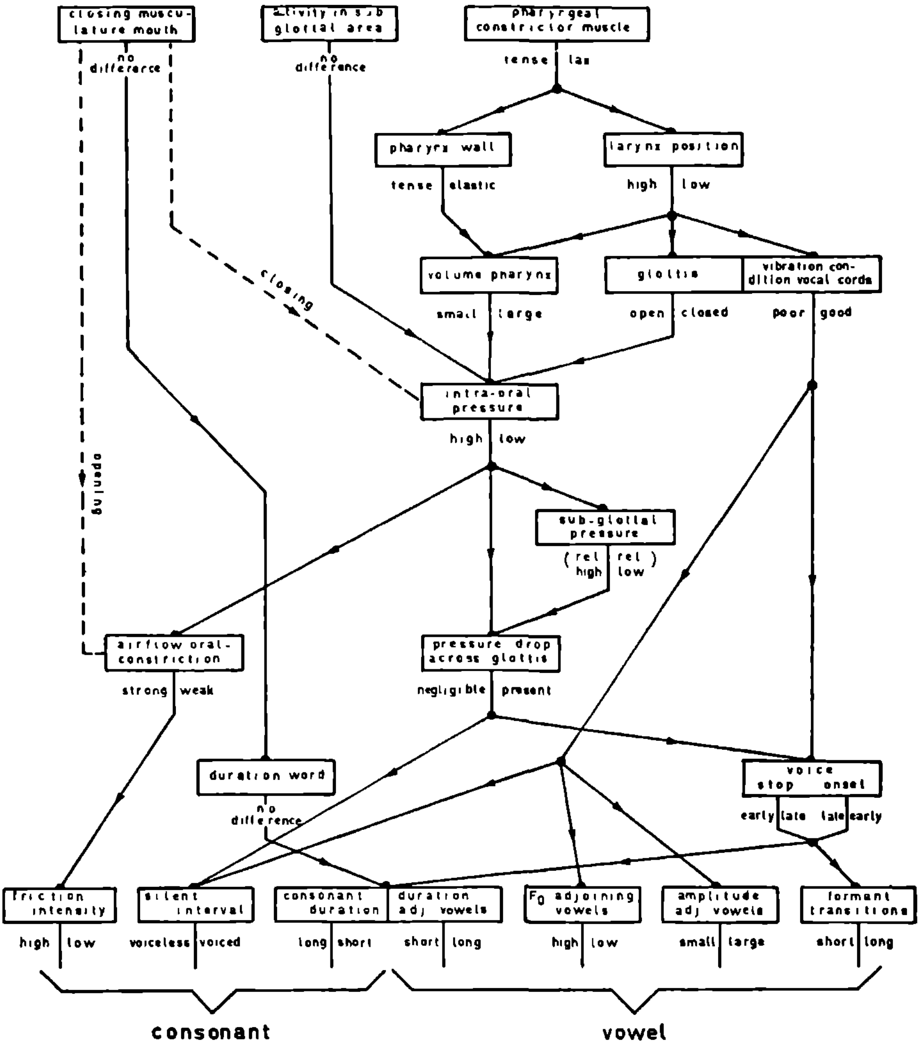


Fig. 18. Schematic representation of the model accounting for the voiced-voiceless distinction in terms of articulatory and acoustic parameters. On the right of the linking lines the voiced situation is indicated, on the left that of the voiceless counterpart.

derivatives, volume of the pharynx and state of the vocal mechanism. The branching lines linking the various boxes are accompanied by texts representing the voiceless situation, left, and voiced situation, right.

The middle part of the diagram (left) is supposed to deal with the air pressure phenomena which are controlled by the events portrayed at the top, as described above. For a proper understanding of this part of the complex it will be related to the total articulatory process.

The air supply needed for phonation derives from the subglottal area. Basing ourselves on Ladefoged (1961) saying that ". . . the muscles regulating the air pressure have to be operated in such a way as to maintain a constant mean background pressure in the lungs . . ." (p. 76), we assume that this part of the mechanism cannot be held responsible for controlling the voiced-voiceless distinction.

Therefore, assuming an equal air supply in both conditions, a modification of the airstream controlling the voiced-voiceless distinction must be looked for in the supraglottal cavities. As soon as the mouth starts closing for the production of consonants, the intra-oral pressure increases, dependent on the volume of the cavities. Volume differences will lead to differences in intra-oral pressure build-up, causing a change in resistance for the upward airstream. This resistance, together with that in the glottis might cause a modification in the subglottal air pressure. This could explain the observed higher subglottal pressure with voiceless consonants (Fischer-Jørgensen, 1963, referring to Roudet); in the diagram this feature is expressed in terms of relatively high *v.* relatively low subglottal pressure. The difference between intra-oral and subglottal pressure accounts for the pressure drop across the glottis.

The moment the mouth opens again, the intra-oral pressure built up during the closure phase is reduced. The resulting airflow will be strong or weak, depending on the difference between the intra- and extra-oral pressure (Isshiki and Ringel, 1964; Subtelny *et al.*, 1966).

The air pressure mechanism, together with the inner condition of the larynx regulates the behaviour of the vocal cords. To account for part of the acoustical effects, *viz.* those bearing on F_0 and amplitude of adjoining vowels, the following considerations are put forward.

Supposing that the vocal cord conditions are identical for voiced and voiceless consonants, and that the pressure difference across the glottis is solely responsible for the mode of vibration of the vocal cords, one would expect a higher F_0 to be correlated with a higher amplitude on aerodynamic grounds. However, our measurements show a high F_0 with a low amplitude of vowels adjoining voiceless consonants and a low F_0 with a high amplitude of those next to voiced ones. Therefore we must assume that the condition of the vocal cords during the time that vowels adjoin voiceless consonants is different from the condition of the vocal cords while vowels adjoin voiced consonants. Hudgins and Stetson (1935) showed that a downward movement of the larynx is correlated with a lowering of F_0, and an upward movement with a rise of F_0. Moreover, it was shown that for the duration of the voiceless consonants the glottis is often more or less opened while with voiced consonants the glottis is in a phonating position. It seems justifiable to conclude that the position of the larynx has an influence on the condition of the vocal cords; this condition determines the F_0 and the amplitude of the adjoining vowels.

Returning to the diagram (Fig. 18) in the bottom row of boxes, the acoustical effects are represented schematically, showing the effect on the consonant (left) and on the vowel (right). Of the vowel effects, F_0 and amplitude have been dealt with in the previous section, whereas the formant transitions received full treatment earlier. The effects on consonant duration and the silent interval were accounted for together with transitions and the intensity of the friction can be seen to be a direct consequence of the amount of airstream.

Further implications of the model

This subsection will deal with additional implications of the model extended to cover phenomena bearing on the occurrence of the voiced-voiceless distinction in various contexts.

When consonants start a speech utterance, air pressure differences accompanying the voiced-voiceless distinction are less clear-cut than when they are in non-initial positions. With initial consonants the build-up of the intra-oral pressure is gradual with voiced as well as with voiceless consonants (Fischer-Jørgensen, 1963) and the maximum pressure does not differ significantly (Lisker, 1965). The system regulating the voiced-voiceless distinction seems to operate similarly under different conditions, such as whispering and normal speaking, consequently a different system in embedded and initial position is not probable. Therefore the explanation of this effect must be looked for in a mechanism other than the one controlling the voice character.

A possible explanation might be found in the build-up of the subglottal pressure at the beginning of a stretch of speech; in this position the build-up of the subglottal pressure is still in progress and has not yet reached its normal value. As the intra-oral pressure depends on the air supply from the subglottal area, it is dependent on the subglottal pressure; consequently the build-up of the intra-oral pressure of initial consonants can neither show a quick rise nor a (relatively) high maximum, but will rather be commensurate with that of the subglottal pressure.

The context in which a consonant or a vowel is spoken can have consequences in duration. The duration of a consonant after a short vowel (/ɑ/ of the word /bɑCə/ in our measurements) is longer than a similar consonant after a long vowel (/a/ in the word /baCə/), with normal speech as well as whisper. The mean difference in the duration of consonants after /a/ and /ɑ/ was 22 msec. ($s = 3.5$ msec.), whereas a mean difference of 91 msec. ($s = 2.5$ msec.) was measured for the vowels; consequently the difference in the consonants cannot be regarded as a simple compensation of differences in the duration of the vowels.

Fischer-Jørgensen (1964) showed that the post-vocalic consonants in monosyllabic words are longer than in disyllabic words (about 40 msec.) and that the vowels in monosyllabic words are shorter than in disyllabic ones (about 10 msec.). She also showed that the place of articulation of the consonants has consequences for the duration of both the consonant and the adjoining vowels for which the length of the articulatory trajectory can probably be held responsible.

As the articulatory movements mould the acoustical signal (Stevens and House, 1956; Fant, 1960; Fujimura, 1961 b,c) it should be possible to explain the acoustical events, such as formant transitions, by relating them to these movements: the closing of the mouth results in a lowering of F_1 and in a transition of F_2 and higher formants to values corresponding to those characteristic of the place of the constriction (loci).

Articulatory movements, correlating with subsequent speech sounds, tend to overlap in time (so-called coarticulation, Öhman, 1963, 1966a, b). This overlap is sometimes more extensive than the time stretch of the adjoining speech sounds and can therefore be held responsible for the large deviation in the values of measured loci of formant transitions (Lehiste and Peterson, 1961) and for the influence the adjoining vowels exert on the spectral composition of the consonants (especially /x/ and /k/, Halle *et al.*, 1957; Jassem, 1961).

Concluding remarks

(1) Although the model presented here was drawn up to account for the difference between voiced and voiceless consonants in Dutch, we have the impression that in its qualitative aspects it is calculated to account for this difference in other languages as well. This observation does not by any means imply that similar quantitative relations are valid in all languages.

(2) A similar approach to the one presented here might well be applied to account for the distinction between nasals and voiced plosives, where nasals can be regarded as plosives pronounced with lowered velum (e.g. Harris *et al.*, 1962).

(3) The theoretical implications of the model are put forward as lending themselves to experimental verification. A start has been made to this end with respect to vocal-cord behaviour by means of transillumination of the glottis, results of which are presented elsewhere (Slis and Damsté, 1967).

References

Abramson, A. S., Lisker, L. and Cooper, F. S. (1965). Laryngeal activity in stop consonants *Status Report Sp. Res., Haskins Labs.* (New York), No. 4, 6.1

Berg, J. W. van den (1953). *Physica van de Stemvorming met Toepassingen.* Dissertation (Groningen).

Berg, J. W. van den (1958). Myoelastic-aerodynamic theory of voice production. *J. Sp. Hrng Res.*, 1, 227.

Berg, J. W. van den, Zantema, J. T. and Doornenbal, P. (1957). On the air-resistance and the Bernoulli effect of the human larynx. *J. acoust. Soc. Amer.*, 29, 626.

BOREL-MAISONNY, S. (1958). Investigations complémentaires sur la nature des phonèmes et leur audition. *J. Fr. Oto-Rhin. Laryng.*, 7, 183.

DAMSTE, P. H. (1958). *Oesophageal Speech after Laryngectomy*. Dissertation (Groningen).

DAMSTE, P. H. (1959). The glosso-pharyngeal press. *Speech*, Oct. 1959, 1.

DAMSTE, P. H. (1961). Cineradiographic observations of velar and glottal movements. *Proc. 4th Int. Congr. Phon. Sci., Helsinki* (The Hague), 147.

DELATTRE, P .C., LIBERMAN, A. M. and COOPER, F. S. (1962). Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. *Stud. Ling.*, 16, 104.

EIJKMAN, L. P. H. (1933). The internal aspect of the larynx in speech. *Arch. néerl. Phon. exp.*, 8-9, 213 (*Proc. 1st Int. Congr. Phon. Sci.*).

FANT, G. (1960). *Acoustic Theory of Speech Production* (The Hague).

FANT, G., LINDBLOM, B. and MARTONY, J. (1963). Spectrograms of Swedish stops. *STL-QPSR, Roy. Inst. Technol.* (Stockholm), No. 3, 1.

FISCHER-JØRGENSEN, E. (1965). The commutation test and its application to phonetic analysis. In *For Roman Jakobson* (The Hague), 140.

FISCHER-JØRGENSEN, E. (1963). Beobachtungen über den Zusammenhang zwischen Stimmhaftigkeit und intra-oralem Luftdruck. *Zeitschr. f. Phon.*, 16, 19.

FISCHER-JØRGENSEN, E. (1964). Sound duration and place of articulation. *Zeitschr. f. Phon.*, 17, 175.

FUJIMURA, O. (1961a). Some synthesis experiments on stop consonants in initial position. *Mass. Inst. Technol., Quart. Progr. Rep.*, No. 61, 153.

FUJIMURA, O. (1961b). Motion picture studies of articulatory movements. *Mass. Inst. Technol., Quart. Progr. Rep.*, No. 62, 197.

FUJIMURA, O. (1961c). Bilabial stop and nasal consonants: a motion picture study and its acoustical implications. *J. Sp. Hrng Res.*, 4, 233.

GOUGEROT, L., FESSARD, A., VALLANCIEN, B. and GREMY, F. (1958). Variation d'impédance laryngée chez un sujet trachéotomisé lors de la dérivation du courant d'air durant un effort phonatoire. *J. Fr. Oto-Rhin.-Laryng.*, 7, 129.

GRAMMONT, M. (1939). *Traité de Phonétique* (Paris).

HALLE, M., HUGHES, G. W., and RADLEY, J. P. A. (1957). Acoustic properties of stop consonants. *J. acoust. Soc. Amer.*, 29, 107.

HARRIS, K. S., SCHVEY, M. M. and LYSAUGHT, G. F. (1962). Component gestures in the production of oral and nasal bilabial stops (Abstract). *J. acoust. Soc. Amer.*, 34, 743.

HAST, M. H. (1962). Subglottic air pressure and neural stimulation in phonation. *Proc. 12th Int. Sp. & Voice Ther. Conf.* (Padua), 34.

HUDGINS, C. V. and STETSON, R. H. (1935). Voicing of consonants by depressing of larynx. *Arch. néerl. Phon. exp.*, 11, 1.

ISSHIKI, N. and RINGEL, R. (1964). Air flow during the production of selected consonants. *J. Sp. Hrng Res.*, 7, 233.

JASSEM, W. (1961). The acoustics of consonants. *Proc. 4th Int. Congr. Phon. Sci., Helsinki* (The Hague), 50.

KOZHEVNIKOV, V. A. and CHISTOVICH, L. A. (1966). *Speech: Articulation and Perception.* Translation of *Rech: Artikulyatsiya i Vospriyatiye* (Moscow-Leningrad), 1965.

LADEFOGED, P. (1961). Subglottal activity during speech. *Proc. 4th Int. Congr. Phon. Sci., Helsinki* (The Hague), 73.

LEHISTE, I. and PETERSON, G. E. (1961). Transitions, glides and diphthongs. *J. acoust. Soc. Amer.*, 33, 268.

LIBERMAN, A. M., DELATTRE, P. C. and COOPER, F. S. (1958). Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, 1, 153.

LISKER, L. (1965). Supraglottal air pressure in the production of English stops. *Stat. Rep. Sp. Res., Haskins Labs.* (New York), No. 4, 3.1.

LISKER, L. and ABRAMSON, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *Word*, 20, 384.

MALÉCOT, A. and PEEBLES, K. (1965). An optical device for recording glottal adduction-abduction during normal speech. *Zeitschr. f. Phon.*, 18, 545.

MALÉCOT, A. (1966). The effectiveness of intra-oral air-pressure-pulse parameters in distinguishing between stop cognates. *Phonetica*, 14, 65.

MARTONY, J. (1962). Some experiments on perceptual cues for Swedish fricatives. *Sp. Commun. Sem.* (Stockholm), D-4, 8.

ÖHMAN, S. E. G. (1962a). On the perception of Swedish consonants in intervocalic position. *Rep. No. 25, Royal Inst. Technol.* (Stockholm).

ÖHMAN, S. E. G. (1962b). Perceptual segments and rate of change of spectrum in connected speech. *Sp. Commun. Sem.* (Stockholm), D-5.

ÖHMAN, S. E. G. (1963). Coarticulation of stops with vowels. *STL-QPSR, Roy. Inst. Technol.* (Stockholm), No. 2, 1.

ÖHMAN, S. E. G. (1965). On the co-ordination of articulatory and phonatory activity in the production of Swedish tonal accents. *STL-QPSR, Roy. Inst. Technol.* (Stockholm), No. 2, 14.

ÖHMAN, S. E. G. (1966a). Co-articulation in VCV utterances: spectrographic measurements. *J. acoust. Soc. Amer.*, 39, 151.

ÖHMAN, S. E. G. (1966b). Perception of segments of VCCV utterances. *J. acoust. Soc. Amer.*, 40, 979.

PERKELL, J. S. (1965a). Studies of the dynamics of speech production. *M.I.T. Quart. Progr. Rep.*, No. 76, 253.

PERKELL, J. S. (1965b). Cineradiographic studies of speech. *Proc. 5th Int. Congr. Acoust.*, A-32.

ROUDET, L. (1910). *Eléments de Phonétique Générale* (Paris).

SIEVERS, E. (1881). *Grundzüge der Phonetik* (Leipzig).

SLIS, I. H. (1966). A model for the distinction between voiceless and voiced consonants. *Ann. Progr. Rep. Percept. Res.* (Eindhoven), No. 1, 40.

SLIS, I. H. (1967). What causes the voiced-voiceless distinction? *Ann. Progr. Rep., Inst. Percept. Res.* (Eindhoven), No. 2, 71.

SLIS, I. H. and COHEN, A. (1969). On the complex of the voiced-voiceless distinction, I. *Language and Speech*, 12, 80.

SLIS, I. H. and DAMSTE, P. H. (1967). Transillumination of the glottis during voiced and voiceless consonants. *Ann. Progr. Rep., Inst. Percept. Res.* (Eindhoven), No. 2, 103.

STETSON, R. H. (1951). *Motor Phonetics*, 2nd ed. (Amsterdam).

STEVENS, K. N. and HOUSE, A. S. (1956). Studies of formant transitions using a vocal tract analog. *J. acoust. Soc. Amer.*, 28, 578.

SUBTELNY, J. D., WORTH, J. H. and SAKUDA, M. (1966). Intraoral pressure and rate of flow during speech. *J. Sp. Hrng Res.*, 9, 498

VALLANCIEN, B. (1958). Le larynx et sa fonction phonatoire. *J. Fr. Oto-Rhin.-Laryng.*, 7, 137.

VENTZOV, A. V. (1966). A mechanism of production of voiced and voiceless inter-vocalic consonants. *Int. Congr. Psychol., Symp. 23: Models of Speech Perception* (Leningrad), 105.

ZENKER, W. (1966). Fragen der Function der äuseren Kehlkopfmuskeln (Rez.). *Zeitschr. f. Phon.*, 19, 262.

CHAPTER 4

ARTICULATORY MEASUREMENTS ON VOICED, VOICELESS AND NASAL CONSONANTS:

A test of a model

I. H. Slis

Reprinted from:

PHONETICA
vol. 21, 1970, 193-210

Articulatory Measurements on Voiced, Voiceless and Nasal Consonants

A Test of a Model

I. H. Slis

Institute for Perception Research, Eindhoven

1. Introduction

In a previous article [Slis and Cohen, 1969] we showed that the difference between voiced and voiceless consonants, linguistically described by one distinctive feature only, is acoustically a very complicated matter. It proved to be possible to find at least eight different acoustical attributes that seem to be related to the voiced-voiceless distinction, viz. (in arbitrary order):

1. acoustical duration of the consonant,
2. acoustical duration of the preceding vowel,
3. duration and spectral extensiveness of the vowel formant transitions,
4. presence or absence of vocal vibrations during the consonant,
5. acoustical duration of the noise burst of plosives,
6. sound pressure of the noise burst of plosives and friction noise of fricatives,
7. sound pressure of the adjoining vowels,
8. peak value of the fundamental frequency of the surrounding vowels and the contour of the fundamental frequency during the following vowel.

Although the acoustical data obtained by our measurements on Dutch consonants and those found in the current literature for other languages as reviewed in Slis and Cohen [1969] showed quantitative differences, qualitatively the same tendencies could be observed.

The main purpose of the present article is to provide more experimental evidence for an articulatory model (see below). Few physiological data concerning the voiced-voiceless distinction in Dutch were available in the literature and from our own experiments. Since the acoustical data on Dutch and the other languages were generally comparable, we assumed that this would likewise be the case with the physiological (articulatory) data found in the literature, concerning (1) intra-oral pressure, (2) degree of opening of the glottis, (3) width of the pharynx, (4) position of the larynx and (5) the airflow through the oral constriction.

To account for these acoustical and articulatory data we constructed an articulatory model for the voiced-voiceless distinction [Slis, 1966, 1967; Slis and Cohen, 1969]. According to this model the voiceless character comes about by tensing the pharyngeal constrictor muscle; with voiced consonants this muscle should remain lax. We hoped to provide in this way a physiological background for the single distinctive feature voiced-voiceless in the linguistic description. The terms tense and lax seem to be more appropriate than voiceless and voiced.

The pharyngeal constrictor muscle, the origin of the voiced-voiceless distinction in our model, forms the back wall of the pharynx. Among a number of connections, one is with the larynx, to which the vocal cords are fixed. We assumed that with voiceless consonants an action of the pharynx constrictor muscle pulls the larynx upward, which causes the vocal cords both to stretch and move apart (abduction). Whether the glottis really opens during Dutch voiceless consonants (as we assumed) had not been proved experimentally. That this is by no means a trivial question is shown by the fact that many investigators assume that the glottis is closed during voiceless plosives [e.g. Grammont, 1939]. Damsté [1959] held a contraction of the pharynx cavity with closed glottis ('pumping action') responsible for the observed rise of intra-oral pressure. Therefore, we investigated by means of transillumination of the glottis whether the vocal cords were apart or adducted; in fact, we found that the glottis was open during voiceless consonants (see below).

1.1 According to our model, both contraction of the pharynx wall and the consequently changed position of the larynx (pulled up) result in a decrease of the volume of the mouth-throat cavity during voiceless consonants. Apart from a 'pumping action' of the pharynx wall, 'inflating action' of the pharynx cavity should occur through the opened glottis by an air-stream from the lungs to the pharynx. As the volume decreases and the glottal resistance is low (open glottis) the air pressure in the mouth-throat cavity rises quickly, approaching the maximum possible pressure within the short time available for the consonant constriction. The ultimate pressure reached should be high compared to that during voiced consonants, pressure increment during voiced consonants being low because of the large volume of the lax throat cavity and the high resistance in the vibrating glottis.

These predictions are based on the theoretical model; they concern a hypothesis that can be tested by measuring the intra-oral pressure during the articulation of words containing voiced and voiceless consonants. In order to rule out a hypothetical possibility that differences in sub-glottal pressure are the origin of the differences in intra-oral pressure, we measured the sub-glottal pressure at the same time. According to our model we should expect the voice character to have no influence on the sub-glottal pressure or, at the most, a slightly higher sub-glottal pressure with voiced than with voiceless consonants because of a higher resistance in the vibrating glottis. Our measurements gave results that confirmed these expectations (see below).

1.2 In order to explain the relation between beginning and end of the vocal vibration and other articulatory events, such as the pressure drop across the glottis and the state of the oral constriction, we again refer to the model. The driving force for the vibration of the vocal cords is supplied by the air-stream through the glottis. If the intra-oral pressure rises, the pressure drop over the glottis (= sub-glottal pressure minus intra-oral pressure) will decrease. As a consequence, the air-stream through the glottis will also decrease. With voiceless consonants this will happen rather quickly; after a short time the air-stream will be too small to keep the vocal cords in vibration. A 'silent interval' (no vibration of the vocal cords) during the voiceless consonant will result. With voiced consonants the pressure drop will remain sufficiently high to keep the vocal cords in vibration (which in fact they do). This continuation of voice during the consonant we will call the 'muffled interval' which is acoustically limited by a sudden decrease and increase of the acoustic environment.

As we assume in our model that the pharynx constrictor muscle is the only originator of the voiced-voiceless distinction, we conclude that there will be no difference in the action of the closing musculature.

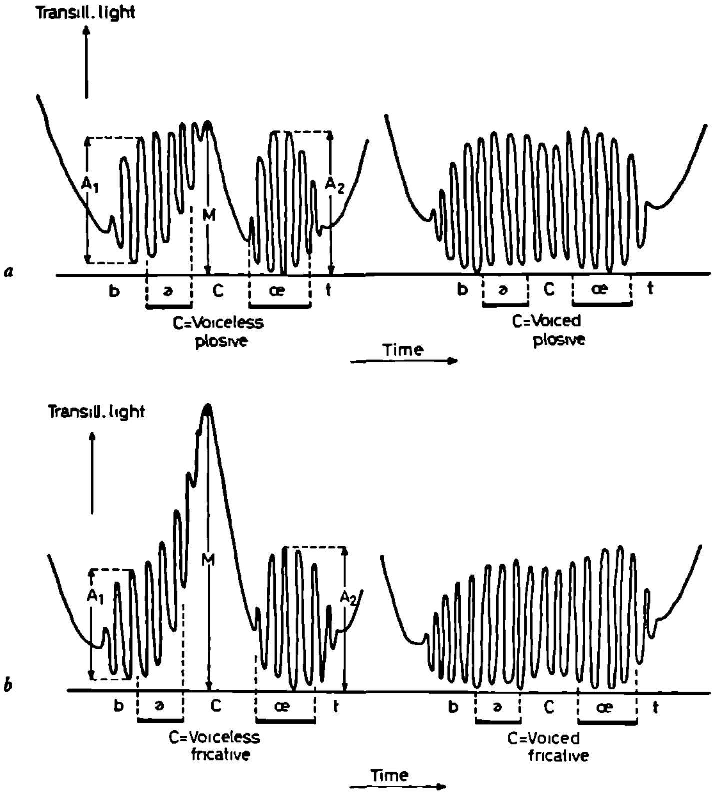


Fig. 1. Schematic drawings of the glottal transillumination curves obtained with: (a) plosives and (b) fricatives in words of the type /beCœt/. A is the amplitude of the envelope of the glottal opening during the adjoining vowels and M is the peak value of the glottal opening during voiceless consonants. The intervals during which (from the acoustical signal) the opening of the oral constriction was evident, are indicated by horizontal bars under the figure. The basis of the figure is a horizontal line through the lowest points in the series of recordings, indicating that the glottis is closed.

Consequently, the duration of the oral constriction will be the same for voiced and voiceless consonants. Historically, our argumentation was the other way round. Since we found that a difference in voice character of an embedded consonant does not influence total word duration, we concluded that we had to look for a muscle or muscle group that is independent of the closing musculature in order to find the origin of the voiced-voiceless distinction.

Acoustically, the muffled (voiced) interval proved to be shorter than the silent (voiceless) interval, which difference was fully compensated by the adjoining vowels. The model predicted equal oral constrictions with voiced and voiceless consonants; this discrepancy between acoustical and hypothetical articulatory events leads to the assumption that the vowel before a voiceless consonant ends acoustically even before the mouth is closed. We attributed the early ending of the voice to a decrease of the pressure drop across the glottis before the actual closing of the mouth. This decrease should be due both to an increase of resistance in the oral constriction during the closing gesture of the mouth, and to a simultaneous opening of the glottis. In order to test this hypothesis we measured closure durations of the lips in words containing /p/ and /b/. The measurements showed that closure durations of the lips during /p/ and /b/ are not equal and, therefore, we had to revise the model (see below).

1.3 By extending the model slightly we were able to include the nasals in our measurements on the assumption that they could be considered as voiced plosives with a lowered velum.

In the following pages we will describe measurements on (a) the behaviour of the glottis, (b) sub-glottal and intra-oral pressure, and (c) closure duration of the lips during words containing voiced and voiceless consonants as well as nasals. Because of a discrepancy between the model and the measurements of the closure durations of the lips, the model had to be slightly modified. Therefore, a modification to the model will be described, based on 'cross talk' between the articulatory commands. Finally, electromyographic measurements will lead to a revised and refined modification.

2. *Measurements*

2.1 Glottal Opening

The degree of opening of the glottis was measured by transilluminating the glottis [a technique described by Sonesson, 1960, and others]. If a beam of light is directed onto the skin of the throat under the larynx, part of the light is transmitted through the wall of the trachea and scattered inside the trachea. When the glottis is open some of this light shines through it, the amount being proportional to the surface of the glottis. A bundle of glass fibres, introduced into the throat through the nose, conducts part of this light to a photomultiplier outside the subject [SLIS and DAMSTÉ, 1967].

One subject pronounced words of the type /bəCœt/ and /baCə/ in which /C/ stands for /p, t, k, f, s, x, b, d, v, z, m, n/. During these pronounciations, measurements were made simultaneously of the amount of light and of the acoustical signal. The results showed that the glottis was open during all voiceless consonants and vibrated during all voiced consonants.

For the voiceless consonants, the beginning and end of the open phase were determined from the recordings by visual criteria. The opening movement of the glottis started in /bəCœt/ words about 70 ms and in /baCə/ words about 40 ms before the acoustical end of the preceding vowel, reached a maximum at about 40 (/bəCœt/ words) or 70 ms (/baCə/ words) after the acoustical end of the preceding vowel and ended at the beginning of the following vowel (fig. 1).

During the open phase of voiceless plosives the mean maximal glottal opening (M in fig. 1a) proved to be equal to the mean maximum amplitude of the envelope of the glottal vibration in the surrounding vowels (A in fig. 1a), thus M = A (table I). M proved to be about 50% higher during voiceless fricatives than during voiceless plosives; A is smaller with fricatives than with plosives (table I). Thus we observed an opening of the glottis with voiceless consonants. The opening movement starts before the end of vocal vibration; this is in accordance with our model. This model predicted an open glottis during voiceless consonants and a continuation of vocal vibration until the pressure drop across the glottis became too small to maintain vibration.

2.2 Intra-Oral and Sub-Glottal Pressure

Thanks to the mediation and cooperation of Dr. DAMSTÉ of the Academic Clinic of the University of Utrecht we were able to use for our purposes the apparatus of the cardiographic department of this hospital, designed to measure blood pressure. The intra-oral pressure was measured directly by means of an open rubber tube which was introduced into the pharynx through the nose and which was connected with a pressure transducer outside the subject. The sub-glottal pressure was measured indirectly by means of a small rubber balloon

Table I. Maximum degree of opening of the glottis and the amplitude of the envelope of glottal vibration during the adjoining vowels, measured by means of transillumination of the glottis. The numbers given are relative mean values in arbitrary units during two series of measurements, viz. /bɔCœt/ and /baCɔ/ words

	Amplitude of glottal opening during preceding vowel A1	Maximum glottal opening consonant M	Amplitude of glottal opening during following vowel A2	Number of measurements n
/bɔCœt/, C = voiceless plosive	19	11	17	9
/bɔCœt/, C = voiceless fricative	14	27	15	8
/baCɔ/, C = voiceless plosive	23	34	10	8
/baCɔ/, C = voiceless fricative	20	46	10	7
Mean of voiceless plosives	21	22	16	17
Mean of voiceless fricatives	17	36	10	15

Table II. Peak values of the sub-glottal pressure (in mm H_2O) during different classes of consonants. The number given is the mean of n measurements

	Normal speech	Shouting	Whispering	Number of measurements, n
Voiceless plosives	115	180	100	6
Voiceless fricatives	110	180	100	6
Voiced plosives	110	195	100	4
Voiced fricatives	125	200	100	4
Glides, nasals, liquids	110	190	100	12
/h/	85	195	65	2

that was introduced into the oesophagus; the balloon was connected, through the nose, with a pressure transducer outside the subject by means of a rubber tube [Slis and Damsté, 1967]. This method has been described in detail by Van den Berg [1956].

The measurements were done again on embedded consonants in words of the type /bəCœt/ and /baCə/, in which /C/ stands for /p, t, k, f, s, x, b, d, v, z, m, n, w, j, l, r, h/. These words were spoken in a normal way, shouted and whispered by one subject. The results showed that peak values of the sub-glottal pressure seem to be independent of the consonant except for /h/ during which a pressure drop could be observed (table II). We may, therefore, conclude that differences in intra-oral pressure (see below) do not have their origin

Table III. Peak values of the intra-oral pressure (in mm H_2O) during different classes of consonants. The number given is the mean of n measurements

	Normal speech	Shouting	Whispering	Number of measurements n
Voiceless plosives	60	130	85	6
Voiceless fricatives	60	125	70	6
Voiced plosives	35	65	70	4
Voiced fricatives	50	95	60	4
Glides, nasals, liquids	20	20	40	12
/h/	5	5	5	2

in differences in sub-glottal pressure; on the contrary, during shouting we even observed a tendency to a higher sub-glottal pressure in voiced consonants than in voiceless consonants, while the intra-oral pressure was lower. We can explain this as being due to a higher resistance in the vibrating glottis (see above).

The intra-oral pressure both in normal speech and in shouting is considerably higher during voiceless than during voiced consonants (table III). In whispered speech these differences are barely observable in our material; the open 'whisper configuration' of the glottis levels the difference between the resistance in an open glottis (voiceless configuration) and in a closed glottis (voiced configuration). The small differences that remain in whispered speech will be due to the contraction of the pharynx. Comparison between normal and whispered speech illustrates the importance of the glottal resistance during normal speech in the build-up of the intra-oral pressure (table III).

In normal speech the intra-oral pressure shows a sudden rise, presumably due to closing of the mouth, followed by a more gradual rise, presumably due to 'inflation' of the pharynx (fig. 2).

The results above seem to fit perfectly into our model with the exception of the sudden rise of the intra-oral pressure; in voiceless consonants we expected that the vocal vibration would stop before the mouth is closed because of an increase in intra-oral pressure during the closing movement of the mouth. In this view, it follows from the gradual closing of the mouth and from the gradual opening of the glottis that the pressure increase should be gradual. The discontinuous jumb in pressure must, therefore, be attributed to a discontinuity in the articulation movement, viz. the closing of the constriction. We

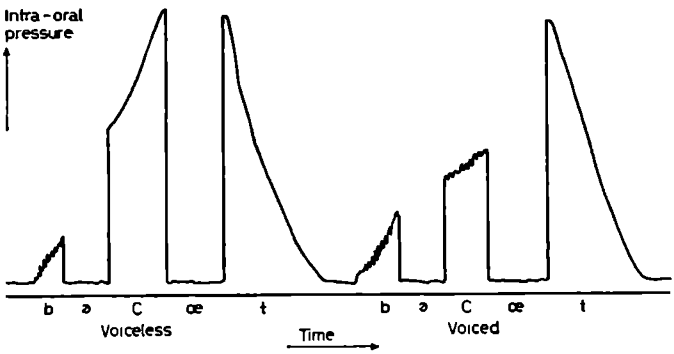


Fig. 2. Schematic drawing of the contour of the intra-oral pressure during words of the type /bəCœt/.

will see that the assumption of an equal duration of closure in voiced and voiceless consonants was wrong (see below).

2.3 Closure Duration of the Lips

The intervals during which the lips are closed in /p/, /b/ and /m/ were measured by means of 'lip contacts' [Slis, 1968]. These lip contacts consisted of small silver strips that were bent over the lips. One of the strips was covered with textile (with a low electrical resistance) to prevent repetitive interruption of the contact at the moment the lips are brought together. Simultaneously with the moments of closing and opening of the lips, the beginning and end of voice activity were recorded; the voice activity was measured by means of a throat microphone.

We will report here measurements on words of the type /bəCœp/, /Cap/, /baCə/ and /b*a*Cə/ in which /C/ stands for /p/, /b/ or /m/ [Slis, 1968]. We found that the interval during which the lips are closed in embedded /p/ is always longer (15–20 ms) than in embedded /b/ and /m/; with initial /p/, /b/ and /m/ no such difference in lip closure duration is observed (table IV).

The total word duration proved to be independent of the voice or nasal character of the embedded consonant. Consequently, differences in duration of lip closure in embedded /p/, /b/ and /m/ are compensated by differences in duration of the other speech sounds; in words of

Table IV. Duration of speech sounds (in ms), based on the intervals the lips are closed or opened. Three series of words were run, viz. (a) /Cap/words, (b) /baCə/- and /bɑCə/words, and (c) /bəCœp/words in which C = /p/, /b/ or /m/. The values placed between brackets are the intervals between the moment of lip opening and the end of voice activity. Since the end of voice activity is more difficult to measure than the lip movements, these values are less accurate than those obtained by lip measurements

(a)			Initial /p, b, m/	Following vowel /a/	Number of measurements n		
/pap/			122	177	90		
/bap/			129	172	90		
/map/			132	163	90		
(b)	**Initial /b/**	**Stressed vowel /a, ɑ/**	**Embedded /p, b, m/**	**Final /ə/**	**Total word duration**	**Number of measurements n**	
/bapə/	130	178	87	(121)	(516)	60	
/babə/	132	197	67	(123)	(519)	60	
/bamə/	136	206	76	(112)	(530)	60	
/bɑpə/	144	113	96	(114)	(467)	60	
/bɑbə/	148	122	72	(126)	(468)	60	
/bɑmə/	149	131	74	(122)	(476)	60	
(c)	**Initial /b/**	**Unstressed /ə/**	**Embedded /p, b, m/**	**Stressed /œ/**	**Final /p/**	**Total word duration**	**Number of measurements n**
/bəpœp/	108	73	94	110	95	479	150
/bəbœp/	119	78	79	110	95	481	150
/bəmœp/	125	76	82	100	94	477	150

the type /bVCə/ (in which V =/a/ or /ɑ/), the differences in lip-closure duration of embedded /C/ are mainly compensated by differences in duration of the vowel; in words of the type /bəCœp/ the compensation mainly takes place in the initial /b/.

The end of the voice activity in an embedded voiceless consonant (/p/) proved to be about 20 ms after the moment of lip closure; this happens with embedded as well as with final /p/ in /bəpœp/ and with embedded /p/ in /bapə/.

We assumed that closure duration would be independent of the voice character of the embedded consonant and that with voiceless consonants the voice would end before the closure of the mouth (see above). The observed longer closure duration in embedded /p/ than in /b/ and /m/ and the fact that the voice of the vowel before /p/ ends after the lip closure are in conflict with these assumptions.

3. Modification of the Model

The fact that the intervals of lip closure differ in embedded /p/, /b/ and /m/ compels us to modify our model. The assumption that the voiced-voiceless distinction does not influence the duration of the lip closure was one of the main arguments in favour of our choice of the pharynx constrictor muscle as the origin of this distinction. Therefore, we have to extend the possible origins to the less specific pharynx musculature in general which has connections with the oral closure musculature.

Apart from the pharynx musculature as a cause of the difference between voiced and voiceless consonants we shall have to look at the closure musculature as an origin of the difference in duration. We are aiming at a common origin for both the differences that are in accordance with our model as well as for those that are not, viz. the differences in closure duration of the mouth. Therefore, in this stage of the experiments we made the following modification to our model [Slis, 1968] which we shall present here in the form of a metaphor:

If we assume that our nervous system is comparable to an electronic machine and that the electric wiring of this machine is not ideal, we might expect 'cross talk' between the 'command lines'. A speculation of the site of cross talk in the nervous system would be the motor cortex because of a topographical overlap of the representation of different speech functions [Luchsinger and Arnold, 1965; p. 434]. The innervation directed to the pharynx musculature with voiceless consonants would 'leak' into the 'command lines' to the closure musculature (lips, tongue), and perhaps other muscles as well, giving rise to stronger commands with voiceless than with voiced consonants. The stronger command to, say, the lips results in a quicker movement and consequently in an earlier closing of the lips with /p/ than with /b/ (or /m/).

The increase due to cross talk should occur simultaneously with the innervation of the pharynx musculature. With the transillumination experiments we observed that in /bəpœt/ the opening of the glottis started about 70 ms before the end of the /ə/ (preceding vowel). It is reasonable to assume that there is a delay of more than 5 ms between the innervation of the pharynx musculature and the observed opening of the glottis due to muscle properties and mechanical inertia of the organs under discussion. Since the duration of the /ə/ is about 75 ms we can be sure that the influence of the voiceless consonant /p/ is already present at the moment the mouth opening of initial /b/ occurs. As a consequence of neural cross talk during the opening movement of the initial /b/ the mouth will open faster with a voiceless consonant following after the /ə/ than with a voiced one following.

With the words /bapə/ and /bɑpə/ the first vowel is much longer than in /bəpoep/; consequently, the influence of the following consonant need not yet be present at the beginning of the vowels /a/ and /ɑ/. Therefore, the speed of lip opening and the closure duration of the lips of initial /b/ can be independent of the voice character of the embedded consonant in /bVCə/ words.

In table V we see that the interval between the lip closing of the initial /b/ and the lip opening of the embedded /p/ is equal; thus, differences in closure duration of the lips are compensated within the first part of the word. The moment of lip opening of the embedded consonant is, therefore, independent of the voice character. Moreover, at the moment the mouth opens after embedded /p/ the glottis has already closed again; consequently, we may safely assume that the influence of the 'feature' voiceless is no longer present.

The embedded nasal /m/ behaves as a voiced plosive /b/ with respect to its influence on the duration of the initial /b/ of the word and to the duration of the open phase of the lips of the preceding vowel (/a/, /ɑ/ or /ə/, see table IV). The moment of lip opening after /m/, however, proves to be delayed compared with /p/ and /b/ as following from the fact that the interval between the lip closing of initial /b/ and the lip opening of embedded /C/ is longer with /m/ than with /p/ or /b/ (table V). This might be due to mechanical reasons; the intra-oral pressure during /p/ and /b/ is considerably higher than

Table V. Duration (in ms) of the interval between the moment of closing of the lips in initial /b/ and the moment of opening the lips in embedded /p/, /b/ or /m/, in words of the types /baCə/, /bɑCə/ and /bəCœp/

	/baCə/ words	/bɑCə/ words	/bəCœp/ words	Mean of three types of words
C = /p/	395	353	274	340
C = /b/	396	342	276	336
C = /m/	418	354	283	352

during /m/ (table III) owing to the leak through the nasal passage. In /p/ and /b/ the high pressure drives the lips apart [Fujimura, 1961] resulting in a fast opening movement, but in /m/ the lower pressure will result in a slightly slower opening of the lips.

4. *Electromyographic Measurements*

In short: we concluded above that a hypothetical cross talk between articulatory commands may account for differences in duration of the individual speech sounds, as measured by opening and closing of the lips, in words differing in voice character only. The strong commands during voiceless consonants leak into other 'command lines', giving rise to stronger innervation and, consequently, faster movement of the articulators.

In order to find more evidence that agrees with our cross talk metaphor we carried out electromyographic measurements on the closing musculature (presumably: depressor labii inferioris and depressor anguli oris) and the opening musculature (orbicularis oris muscle) of the mouth. The lip closing activity was measured by placing two electrodes on the superior edge of the vermillion border of the lips, the lip opening activity by placing two electrodes 2 cm below the corner of the lips. The electrodes were surface electrodes of the type described by Cooper [1964]. They were metal cups with a diameter of 6 mm that were sucked onto the skin by having the air pumped out of them. The EMG activity was recorded on a tape recorder; most of the artefacts of incidental shifting of the electrodes were removed by this 'high pass filtering' of the EMG signal [Fromkin and Ladefoged, 1966]. The output of the tape recorder was rectified; the rectified signal passed an RC system with a time constant of 20 ms. Thus, 'activity contours' were obtained which were recorded on a UV-recorder (Honeywell, visicorder), simultaneously with the output of a throat microphone and the output of a microphone that was placed in front of the mouth.

Measurements of both opening and closing activity of the lips were made during the articulation of words of the type /bɔCœp/ and of only the closing activity during the artilulation of words of the type /aCa/, /baCə/ and /Ca/, in which /C/ stands for /p/, /b/ or /m/; in the series in which the opening activity was measured with words of

Table VI. EMG peak values and durations of EMG activity of lip closing of /p/, /b/ and /m/. Peak values are relative to the EMG peak of /p/ (per definition 100); durations are given in ms measured from the start of the EMG activity to the end of EMG activity. The moment of start and end are evaluated by eye. 40 measurements were done with words of the type /bəCœp/, 10 with words of the type /aCa/ for each subject, 10 with words of the type /baCə) and 30 measurements with /Ca/ words

	Subj. IS /bəCœp/ peak	dur.	Subj. SN /aCa/ peak	dur.	Subj. IS /aCa/ peak	dur.	Subj. IS /baCə/ peak	dur.	Subj. IS /Ca/ peak	dur.	Mean values peak	dur.
C = /p/	100	145	100	171	100	149	100	169	100	–	100	157
C = /b/	89	136	78	157	89	144	73	141	93	–	85	144
C = /m/	81	136	77	150	92	127	78	135	79	–	81	136

Table VII. EMG values of lip opening of the embedded /C/ in words of the type /bəCœp/ in which /C/ stands for /p/, /b/ or /m/. The peak values are relative to the peak of /p/ (100); durations are measured in ms. Each value is the mean of 30 measurements

	Peak	duration
C = /p/	100	125
C = /b/	86	113
C = /m/	89	117

the type /bəCœp/, words with embedded /t/, /d/ and /n/ were included besides /p/, /b/ and /m/.

On the basis of the difference in closure duration of the lips we expect to find a higher innervation of the closing muscles with /p/ than with /b/ and /m/. A higher EMG activity of the opening movement of initial /b/ is expected as well if a voiceless consonant follows than if this consonant is voiceless or nasal.

The EMG lip-closing activity of embedded /p/ proved to be significantly higher than that of embedded /b/ and /m/ (15–20%, table VI). The EMG lip-opening activity of embedded /p/ proved to be about 15% higher than that of embedded /b/ or /m/ (table VII). Not enough measurements were made to establish a significant difference between /b/ and /m/ with respect to the EMG activity of closing and opening of the lips (tables VI and VII).

Differences in peak values of EMG activity of the opening and closing movement of the lips in initial /b/ of words of the type /bəCœp/ proved to be very small, although a tendency to a higher activity

Table VIII. EMG values of lip opening of initial /b/ in words of the type /bɔCœp/ in which /C/ is /p/, /t/, /b/, /d/, /m/ or /n/. The peak values are relative to the value of embedded /p/ in /bɔpœp/ (100); durations are in ms

	Peak	dur.		Peak	dur.		Peak	dur.
emb. C = /p/	68	96	emb. C = /t/	87	130	mean emb. C = /p, t/	78	112
emb. C = /b/	65	99	emb. C = /d/	79	128	mean emb. C = /b, d/	72	114
emb. C = /m/	68	95	emb. C = /n/	85	125	mean emb. C = /m, n/	76	110

Table IX. EMG values of lip closing of initial /b/ in words of the type /bɔCœp/ in which /C/ is /p/, /b/ or /m/. The peak values are relative to the peak of lip closing of embedded /p/ in /bɔpœp/; durations are in ms

	Peak	duration
emb.C = /p/	71	127
emb.C = /b/	71	127
emb.C = /m/	65	136

could be observed when a voiceless plosive follows in an embedded position than when a voiced plosive or nasal consonant follows (table VIII and IX).

Although some of the differences in EMG activity seem to occur as expected, we nevertheless believe the magnitude of these EMG differences to be too small to account for the differences in the time domain. Moreover, we find differences in the EMG activity of the opening movement of /p/, /b/ and /m/ which we did not expect. Therefore, we conclude that our cross talk metaphor is too simple.

Furthermore, measurements of the interval between the highest EMG activity and beginning or end of the consonants, as indicated by the acoustical signal, show a fixed relation between closing activity of the lips and the beginning of the consonant and between opening activity of the lips and the end of the consonant. This relation is independent of the character of the consonant; the top of the closing activity occurs 20 ms before the beginning of the consonant (closing of the mouth) and the top of the opening activity occurs about 40 ms before the acoustic end of the consonant (opening of the mouth).

Consequently, the difference in duration of the lip closures and the lip openings are already present in the timing of the EMG activities.

Although the comparison with cross talk in an electronic machine proved to be insufficient, it is reasonable to assume that other kinds of interaction between the commands take place which may explain a difference in the timing of the combination /bə/ in /bəCœp/ dependent on the voice character of the embedded /C/. We think it probable that such an interaction occurs on a high neurological level; we surmise that an interaction between motor cortex and the formatio reticularis could translate 'stronger commands' into 'earlier commands', the latter having the reputation of exerting an influence on temporal processes [e.g. WOOLDRIDGE, 1963].

In general we see from the EMG measurements that the EMG activity is greater with /p/ than with /b/ and /m/. This is observable in the consonant itself and to a lesser extent in the preceding (initial) consonant /b/ in words of the type /bəCœp/. The EMG activity is measured both in peak values and in duration of the activity.

A second conclusion drawn from the EMG measurements is that the EMG peak occurs before the resulting movement. The interval between EMG peak and its articulatory result tends to be constant.

5. *Conclusion*

Measurements of the opening of the glottis and of the subglottal and intra-oral pressure showed that the general purport of the model, which we proposed earlier, for the voiced-voiceless distinction of consonants can be maintained; the glottis proved to be open during voiceless consonants and the intra-oral pressure proved to be higher during voiceless than during voiced consonants. However, a few details of the model had to be modified; because of differences in duration of the closed interval of the mouth between /p/ and /b/ we cannot restrict the origin of the voiced-voiceless distinction to the pharyngeal constrictor muscle, but we have to regard other muscles of the pharynx wall as possible contributors to this distinction as well. It is impossible to explain the difference in closure duration of the lips from the pharynx musculature. A simple hypothesis based on cross talk of the strong articulatory command to the pharynx musculature into the 'command lines' of the other articulators proved to be insufficient, since the timing of the EMG activity of the lips was found to be different with voiceless and voiced plosive /p/ and /b/; therefore, we have to assume that some other kind of interaction takes place between the successive consonants on a high neurological level.

Measurements made with words containing /m/ compared to those with /p/ and /b/ showed that /m/ has a great resemblance to /b/ with respect to glottal behaviour and with respect to influence on the time structure of the context. Therefore, it is plausible to assume that only a single difference exists between /m/ and /b/, such as the opening or closure of the velum.

Acknowledgement

The author wishes to thank Dr. P. H. DAMSTÉ (University of Utrecht) and Ir. L. F. WILLEMS (IPO) for their cooperation in solving physiological and electronic problems.

Summary

In a previous article a model for the voiced-voiceless distinction was presented. In order to provide a broader base for this model additional measurements were carried out on the glottal opening (by transillumination of the glottis), on the sub-glottal and intra-oral pressures, on closure duration of the lips and on EMG activity of the closing and opening movement of the lips. In general, the model could be shown to make correct predictions, although a modification was necessary in the form of the assumption of neural interaction at a high neurological level. Comparison of words with /p/, /b/ and /m/ showed that nasals behave similarly to plosives in their influence on the time structure of the context; the model can be extended to nasals by assuming a difference in velar action only (e.g. between /m/ and /b/).

Zusammenfassung

Artikulationsmessungen stimmhafter, stimmloser und nasaler Konsonanten. Test eines Modells

In einem vorherigen Artikel wurde ein Modell über den Unterschied zwischen stimmhaften und stimmlosen Konsonanten angeboten. Eine festere Grundlage für dieses Modell wird geschaffen mittels additioneller Messungen 1. der glottalen Öffnung (Transillumination der Glottis), 2. des subglottalen und intraoralen Luftdrucks, 3. der Verschlußdauer der Lippen und 4. der EMG-Aktivität der Schließungs- und Öffnungsbewegung der Lippen. Im großen und ganzen konnte gezeigt werden, daß mit dem Modell die richtigen Prognosen zu machen sind, obschon eine Modifikation notwendig war in Gestalt einer Annahme von neuralen Einflüssen auf hoher Ebene. Messungen an Wörtern mit /p/, /b/ und /m/ zeigten, daß der Einfluß der nasalen Konsonanten auf die Zeitstruktur im Kontext vergleichbar ist mit dem der Verschlußlaute. Durch die Annahme einer unterschiedlichen Aktion des Velums kann das Modell auf die nasalen Konsonanten ausgedehnt werden.

Résumé

Mesurages de l'articulation de consonnes sonores, sourdes et nasales. Test d'un modèle

Dans un article précédent un modèle a été présenté représentant la distinction entre des consonnes sonores et sourdes. Afin de fournir à ce modèle une base plus solide, nous avons fait des mesurages supplémentaires 1. de l'ouverture glottale (avec transillumination de la glotte), 2. de la pression sub-glottale et intra-buccale, 3. de la durée de la fermeture des lèvres et 4. de l'activité électromyographique des lèvres. En général il était possible de faire des prédictions correctes en se servant de ce modèle, bien qu'il fût nécessaire d'introduire une modification, à savoir d'une interaction nerveuse à un niveau assez élevé. Comparaison des mots contenant /p/, /b/ et /m/ intervocaliques a montré que la structure temporelle du contexte était influencée d'une façon parallèle; en admettant un rôle different du voile du palais on pourrait ajouter une consonne nasale au modèle présenté.

References

Van den Berg, J.: Direct and indirect determination of the mean subglottic pressure. Folia phoniat. *8:* 1–24 (1956).

Cooper, F. S.: Instrumentation methods for research in phonetics. Proc. 5th int. Congr. phon. Sci., pp. 142–171 (Karger, Basel/New York 1964).

Damsté, P. H.: The glosso-pharyngeal press. Speech oct., pp. 1–7 (1959).

Fromkin, V. and Ladefoged, P.: Electromyography in speech research. Phonetica *15:* 217–242 (1966).

Fujimura, O.: Bilabial stop and nasal consonants. A motion picture study and its acoustical implications. J. Speech and Hearing Res. *4:* 233–247 (1961).

Luchsinger, R. and Arnold, G. E.: Voice-speech-language (Wadsworth, Belmont 1965).

Slis, I. H.: A model for the distinction between voiceless and voiced consonants, IPO-APR I, pp. 40–44 (1966).

Slis, I. H.: What causes the voiced-voiceless distinction? IPO-APR II, pp. 71–76 (1967).

Slis, I. H. and Damsté, P. H.: Transillumination of the glottis during voiced and voiceless consonants. IPO-APR II, pp. 103–109 (1967).

Slis, I. H.: Experiments on consonant duration related to the time structure of isolated words. IPO-APR III, pp. 71–80 (1968).

Slis, I. H. and Cohen, A.: On the complex regulating the voiced-voiceless distinction. I and II. Language Speech *12:* 80–102 and 137–155 (1969).

Sonesson, B.: On the anatomy and vibratory pattern of the human vocal folds. Acta oto-laryng., Suppl. 156 (1960).

Wooldridge, D. E.: The machinery of the brain (McGraw Hill, New York 1963).

Author's address: I. H. Slis, Instituut voor Perceptie Onderzoek, Insulindelaan 2, *Eindhoven* (The Netherlands).

CHAPTER 5

ARTICULATORY EFFORT AND ITS DURATIONAL AND ELECTROMYOGRAPHIC CORRELATES

I. H. Slis

Reprinted from:

PHONETICA
Vol. 23, 1971, 171-188

Articulatory Effort and its Durational and Electromyographic Correlates

I. H. Slis

Instituut voor Perceptie Onderzoek, Eindhoven

Introduction

In this report we will deal with some manifestations of articulatory effort. The notion of articulatory effort in phonetic terminology seems to be based largely on an intuitive approach to speech. We will consider 4 possible cases of more vs. less articulatory effort.

a) Voiceless consonants are felt to be pronounced with more articulatory effort than voiced ones, as appears from the terms 'tense' and 'fortis' for voiceless and 'lax' and 'lenis' for voiced consonants.

b) Stressed syllables are most probably produced with more articulatory effort than unstressed ones; the same may hold for emphatic as opposed to non-emphatic speech.

c) The closing movement of consonants after short vowels may involve more effort than after long vowels, as seems to be indicated in German by the opposition of 'scharf' to 'weich geschnittenen' vowels.

d) In the transition from a consonant to a vowel, the opening of the mouth may involve more effort with a long vowel than with a short vowel. At least this is suggested by the idea that long vowels are 'tense' and short vowels are 'lax'.

We have investigated these 4 cases in which there seems to be an introspectively sensed opposition between more and less articulatory effort, by measuring both consonant durations and the electromyographical activity of opening and closing movements. In all tests the consonants were bilabial, which made it possible to measure consonant durations with lip contacts and the electromyographic activity with surface electrodes.

The intervals during which the lips were closed were measured by means of lip contacts [Slis, 1968]. These contacts consist of small silver strips (0.3 mm thick, 6 mm wide and 40 mm long) (fig. 1) which are bent round the middle of the lips. One strip was covered with material (with a low electrical resistance) to prevent repetitive interruption of the contact when the lips were brought together.

In the tests, stimulus words were presented in a random order to the subjects on a visual display. After presentation of the stimulus word the subject heard a click through headphones. This click indicated the start of a time interval of 2 or 4 sec during which the moments of closing and opening of the lips were recorded. The subjects were asked

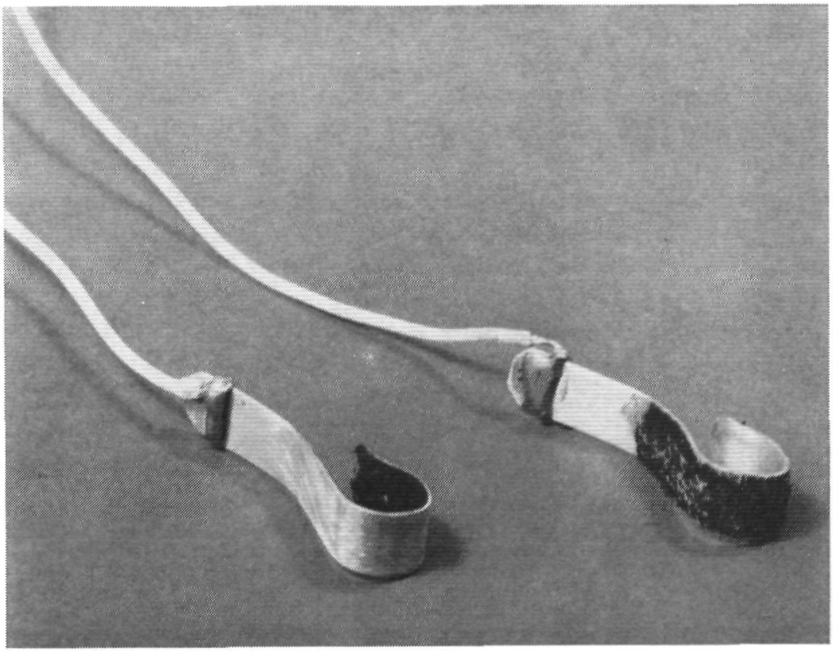


Fig. 1. The lip contacts as used in the experiments. These contacts consist of silver strips (0.3 mm thick, 6 mm wide, 40 mm long) bent so as to fit round the middle of the lips. The strip on the right was covered with material (with a low electrical resistance) to prevent repetitive interruption of the contact.

to say the word they saw on the visual display within this time interval. Some of the results obtained in this way have been published elsewhere [SLIS, 1968; NOOTEBOOM and SLIS, 1969; SLIS, 1970].

1. Voiceless vs. Voiced Consonants

We compared the durations of lip closure and of lip opening for words differing only in the embedded consonant (/p/ vs. /b/) in the word pairs /bapə/ and /babə/, /bɑpə/ and /bɑbə/, and /bəpœp/ and /bəbœp/. The 1st 2 pairs were spoken by 1 subject, the 3rd pair by 2 subjects. The results are given in table I. In column 3 of that table we observe that the closure duration of embedded voiceless /p/ is 15–24 msec longer than that of embedded voiced /b/. Since the total durations of /bVp/ and of /bVb/ are equal (column 4), we conclude that the observed differences between /p/ and /b/ are compensated for in the other phonemes involved.

Columns (1) and (2) indicate that this compensation occurs mainly in the vowel in the case of /bapə/ vs. /babə/ and mainly in the initial consonant in the case of /bəpœp/ vs. /bəbœp/.

2. Stressed vs. Unstressed Syllables

Measurements were carried out on utterances of the type /pVpVpVp/ in which V could be /a/ or /ɑ/. Words with /a/ and /ɑ/ were spoken in

Table I. Duration of speech sounds (msec), based on time for which the lips were closed and open. The series of words of the type (bVCɔ/ were spoken by 1 subject, the series /bəCœp/ by 2 subjects

	Duration of closure of initial /b/ msec (8)	Duration of opening of embedded /a, ɑ, ə/ msec (8)	Duration of closure of embedded /p/ or /b/ msec (8)	Duration of interval /bVp/ or /bVb/ msec (13)	Number of measurements
/bapɔ/	130	178	87	395	60
/babɔ/	132	197	67	396	60
/bɑpɔ/	144	113	96	353	60
/bɑbɔ/	148	122	72	342	60
/bəpœp/	108	73	94	275	150
/bəbœp/	119	78	79	276	150

The value in brackets at the head of a column is an estimated minimum difference in msec that would be significant at a 1% probability level.

separate series. Stress was placed on one of the 3 syllables. Table II compares the durations of lip closure and lip opening. If the syllable was stressed, the closure duration of syllable initial /p/ was 12–24 msec longer than in unstressed syllables. The effects of stress on vowel duration have been discussed elsewhere in detail [NOOTEBOOM and SLIS, 1969; NOOTEBOOM and SLIS, subm. for publ.]; here it may suffice to say that a categorical distinction between long /a/ and short /*a*/ is expressed more clearly by durational differences in the stressed situation.

We are also interested here in possible compensation effects for the differences in consonant duration, particularly in the preceding syllable. In words containing the vowel /a/ we observe that the vowel of the unstressed 1st syllable is shorter with a stressed second syllable following than with an unstressed 2nd syllable following (compare: /papápap/ with /papapáp/).

A comparison of the time structure of the 2nd unstressed syllable shows no significant influence of stress of the following third syllable on vowel duration, but the initial /p/ of the 2nd syllable is significantly shorter if a stressed third syllable follows than when an unstressed third syllable follows (compare: (papapáp/ with /pápapap/). In words with the vowel /*a*/ comparable differences in vowel durations of the 1st syllable and in /p/ durations of the 2nd syllable tend to show up. These differences, however, are not significant.

Table II Duration of the speech sounds (msec), based on the time for which the lips were closed or open, of words of the type /pVpVpVp/ with /V/ = /a/ or /*a*/ and with stress on the 1st, 2nd and 3rd syllable. The duration of the stressed syllabes is underlined. The words were spoken by 2 subjects, each word 20 times. The estimated minimum difference (msec) that would be significant at a 1% level is 10 msec

	Closure /p/	Opening /V/	Closure /p/	Opening /V/	Closure /p/	Opening /V/	Closure /p/
Words with /a/							
/pápapap/	104	132	68	58	70	133	75
/papapap/	84	80	81	140	77	136	77
/papapáp/	83	94	57	60	86	144	84
Words with /*a*/							
/*pápapap*/	112	80	63	60	64	87	78
/*papápap*/	98	73	75	89	63	94	73
/*papapáp*/	94	78	58	62	87	90	80

We may conclude that differences in /p/ duration due to stress on the 2nd syllable are partly compensated for within the preceding 1st syllable in words with the vowel /a/. In words with the vowel /ɑ/ a non-significant tendency in the same sense can be observed. A restriction must be made with regard to the shorter duration of /p/ in the unstressed 2nd syllable if a stressed third syllable follows; this difference might be due to differences in stress of the preceding 1st syllable as well.

3. /p/ After Short and Long Vowels

The closure duration of /p/ after short vowels was compared with that after long vowels in words of the type /bVpə/ in which V could be either the short vowel /ɪ/, /œ/, /ɔ/ or /ɑ/ or the long vowel /e/, /ø/, /o/ or /a/. As we found earlier [Slis and Cohen, 1969], the duration of /p/ after a short vowel is somewhat longer than after a long vowel. In the present experiments this difference amounts to 9–19 msec (table III, column 3). The differences in duration between stressed long and short vowels caused by the phonematic opposition in length is considerably greater than the inverse differences found in the following /p/. Although it seems probably that the same kind of compensation rule holds as in the previous 2 comparisons, we cannot conclude firmly whether or not an exchange of part of the /p/ and part of the vowel takes place, since phonematic differences in vowel length interfere.

4. /b/ Before Long and Short Vowels

The initial /b/ of /bVpə/ words tends to be slightly shorter (5–14 msec) when a long vowel follows than when a short one follows (table III). Since differences in vowel duration caused by differences in vowel length category are much larger, we cannot conclude firmly whether or not the differences in duration of initial /b/ are compensated for in the following vowel.

5. Survey of Time Measurements

Surveying the results of our durational measurements we see that in the cases where on terminological grounds we hypothesized differences in effort the following systematic changes in the time structure of the words can be observed:

1. Voiceless /p/ is longer than voiced /b/ (about 20 msec). This difference in duration is partly compensated for in the preceding speech sounds.

Table III. Duration of the speech sounds (msec) based on the time for which the lips were closed or open. Each word was spoken 60 times by one subject. The words were spoken in the same series, except for /a/ and /ɑ/. An estimated minimum difference (msec) that would be significant at a 1 % level is 8 msec

	Short vowel				Long vowel		
	closure /b/	opening /V/	closure /p/		closure /b/	opening /V/	closure /p/
/bepə/	128	196	97	/bIpə/	133	110	106
/bøpə/	122	187	67	/bœpə/	128	125	86
/bopə/	125	181	76	/pɔpə/	130	126	92
/bapə/	130	178	87	/bɑpə/	144	113	96

2. /p/ in a stressed syllable is longer than /p/ in an unstressed syllable (about 17 msec). Tendencies towards partial compensation in the preceding speech sounds can be observed.

3. /p/ after short vowels is longer than /p/ after long vowels (about 14 msec). No statements can be made with respect to a possible compensation in the preceding vowel for these differences in the duration of /p/.

4. /b/ before a long vowel is shorter than /b/ before short vowel (about 8 msec). No statement can be made about a possible compensation in preceding or following speech sounds.

The rule behind this phenomenon might be that the moment of lip closure is advanced in time if effort is increased. This will be elaborated further in the discussion.

II. Electromyographic Measurements; Methods

The electromyographic closing activity was measured by putting two electrodes on the edge of the upper lip, thus picking up the activity of the orbicularis oris muscle. The electrodes were surface electrodes of the type described by Cooper [1964]. They were metal cups with a diameter of 6 mm held against the skin by suction. The amplitude envelope of the electromyographic activity was recorded on a UV-oscillograph [Slis, 1970]. The acoustic output of the mouth was recorded simultaneously. From these recordings we were able to define the time relation between electromyographic activity and moments of closing (sudden drop of the amplitude of the acoustic signal) and opening (noise burst of the plosive) of the mouth. In order to check whether the lip-opening muscles are more active in long vowels than in short ones we also measured EMG activity of muscles which we assumed open the lips. We found these muscles empirically by trying different electrode locations. Maximum activity coupled to opening was found at a spot 2 cm below the corner of the mouth. At this point we presumably measured the activity of the depressor Labii inferioris muscle and of the depressor Anguli oris muscle. These measurements are less reliable than those on the closing muscles, since movements of the lower jaw caused some artefacts.

The amplitude envelope of both the closing and opening activity was obtained by high-pass 'filtering' of the electromyographic signals by means of a tape recording. This eliminated most of the artefacts due to shifting of the electrodes [Fromkin and Ladefoged, 1966]. The output of the tape recorder was rectified and passed through an RC filter with a time constant of 20 msec. The muscle activity was measured in terms of the maximum of the amplitude envelope, and the results were subsequently normalized by giving the value 100 to the lowest mean value of each group to be compared, viz. the lip-closing activities of embedded /b/ in /bəbœp/, of embedded /p/ in /pVpV/ and of embedded /p/ in /bVpə/, where /V/ is a long vowel (table IV), and the lip-opening activity of initial /b/ in /bVpə/, where /V/ is a short vowel (table V).

Since we worked with an RC system with a certain inertia, the differences obtained probably correspond not only with differences in activity level but also with differences in duration of the muscle activity. Whether the greater effort is expressed in a more intense innervation or in a longer activation of the muscle did not seem a relevant question at this stage of the experiments, since in both cases the result is an increase in the total amount of activity related to one muscle command.

In order to compare the EMG with the articulatory consequence of muscle activity, we have to select a certain time criterion for the EMG.

We chose for this the moment of maximum EMG activity (the EMG 'peak'), a moment which is, however, influenced to some extend by the RC time constant of 20 msec. Theoretically, the moment at which the EMG activity begins would have been preferable, but in actual practice this is difficult to define. In fact, we did not find substantial differences in rise times of the EMG amplitudes in the initial consonants in which no such difficulties obtain, and this made us confident that the EMG peak suited our purposes.

The moments of lip closing and opening could not be measured with lip contacts since this procedure interferes with the electromyographic measurements. The moment of lip closure was therefore defined as the moment with the most rapid decay of the amplitude envelope of the acoustic signal. The moment of lip opening was taken to be the moment of the noise burst in the oscillographic recordings.

1. Results of the Electromyographic Measurements

The results of the EMG measurements show that at a significance level of 1%:

1. The closing activity of embedded voiceless /p/ is 12% higher than that of embedded voiced /b/ (table IVa).

2. The closing activity of /p/ in the stressed 2nd syllable /pa/ is 16% higher than in the unstressed syllable 2nd /pə/ and the closing activity of embedded /p/ in emphatically spoken words is 20% higher than in normally spoken words (table IVb).

3. The closing activity of /p/ after short vowels is 5 to 10% higher than after long vowels (table IVc).

4. The opening activity of /b/ before long vowels does not differ significantly from that before short vowels. However, the activity of /b/ before long /a/ proved to be significantly higher than before short /ɑ/ (table V).

With regard to the time intervals between the EMG peak and the moment of lip closing or opening, it can be observed that these differ less than 10 msec in all the different oppositions discussed except for 2 pairs, viz. lip closing in /bepə/ vs. /bɪpə/ and lip opening in /bœpə/ vs. /bøpə/ (table IV and V).

III. Discussion of the Results

The measurements of the duration of closure of the lips during labial plosives showed lengthening with increasing effort (1) of the consonant (/p/ vs. /b/), (2) of the syllable of which the consonant is part (stressed vs. unstressed syllables) or (3) of the closing movement of the preceding vowel (short vs. long vowels). A shortening of the closure duration with increasing effort of the following vowel occurred (long vs. short vowels). Differences in closure duration may be caused by an ad-

Table IV. Interval between the peak of the EMG envelope and the moment of lip closing in msec for embedded consonants (column 1), the normalized peak value of the EMG activity (column 2) and the number of measurements (column 3)

	Interval between EMG peak and lip clos. msec	Relative EMG peak amplitude	Number of measurements		Interval between EMG peak and lip clos. msec	Relative EMG peak amplitude	Number of measurements
a)							
/bəbœp/	17	100	40	/bɔpœp/	19	112	40
/pəpa/norm.	8	109	10	/pəpa/emph.	5	128	10
/pape/norm.	24	91	10	/papə/emph.	24	112	10
Mean normal	16	100		Mean emphat.	15	120	
c)							
/bepə/	29	97	30	/bɪpə/	40	107	30
/bøpə/	28	101	30	/bœpə/	29	106	30
/bopə/	15	103	30	/pɔpə/	23	111	30
/bapə/	18	98	30	/pɑpə/	27	105	30
Mean long	22	100		Mean short	30	107	

These measurements concern: (a) the opposition /b/ vs. /p/, (b) normal speech vs. emphatic speech and (c) /p/ after long vowels vs. /p/ after short vowels. All differences between corresponding EMG activities in the left-hand and in the right-hand set of columns are significant at an estimated level of 1 %.

Table V. Interval between the peak of the EMG activity and the moment of lip opening (msec) for initial /b/, the mean peak of the amplitude envelope and the number of measurements

	Interval between EMG peak and lip opening msec	Relative EMG peak	Number of measurements		Interval between EMG peak and lip opening msec	Relative EMG peak	Number of measurements
/bɪpə/	33	130	30	/bepə/	27	134	30
/bœpə/	36	77	30	/bøpə/	50	80	30
/bɔpə/	36	66	30	/bopə/	43	55	30
/bɑpə/	27	127	30	/bapə/	24	142	30
Mean short	33	100		Mean long	36	103	

These measurements concern: the opposition between /b/ before long vs. /b/ before short vowels. The EMG peaks are relative to the mean peak of the short vowels.

vancement of the lip closure or a delay of the lip opening. These differences in timing may be compensated for in the surrounding speech sounds in such a way that the overall time structure remains unchanged. In this case it seems reasonable to suppose a 'low level' origin for the changes in time structure. If no compensation occurs it seems evident that the differences have been programmed on a relatively high level of speech production.

In 2 cases we found total or partial compensation of the differences of closure duration in the preceding vowel, viz. /p/ vs. /b/ and /p/ in stressed vs. /p/ in unstressed syllables; in the 2 other oppositions the compensation is not demonstrable, since the oppositions concern the 2 length categories of the vowel, but it may very well take place. This leads to the postulation of an implementation rule, on a low level of speech production, stating that increasing effort leads to an advancement of lip closure (/p/ vs. /b/, /p/ in stressed vs. /p/ in unstressed syllables and /p/ after short vs. /p/ after long vowels), or lip opening (/b/ before long vs. /b/ before short vowels).

The electromyographic measurements showed that in all lip-closing movements a higher activity could be observed in all cases where a higher effort was assumed. In lip-opening movements a higher activity could be observed with the /a/ vs. /ɑ/ opposition only. With the other 3 vowel oppositions no significant results were found. During the measurement of opening activity the electrodes did not adhere to the skin as they ought to have done since the lower jaw moved too much during speech.

A larger standard deviation therefore occurred with lip-opening measurements than with lip-closure measurements. Better measurements will be needed to obtain trustworthy results on lip-opening activity. We nevertheless think it justified to assume higher electromyographic activity in the 4 effort oppositions discussed.

1. The Effort Implementation Rule

Two different possible explanations may be offered for the observed shift in moment of lip closure. One is that a stronger command results in a faster closing movement (fig. 2a). The alternative is that the command itself is shifted in time (fig. 2b). This is in accordance with a suggestion of LINDBLOM [1963], based on spectrographic measurements of vowels, that the speed of the articulators is limited and that therefore greater effort could not change the speed of the articulators ap-

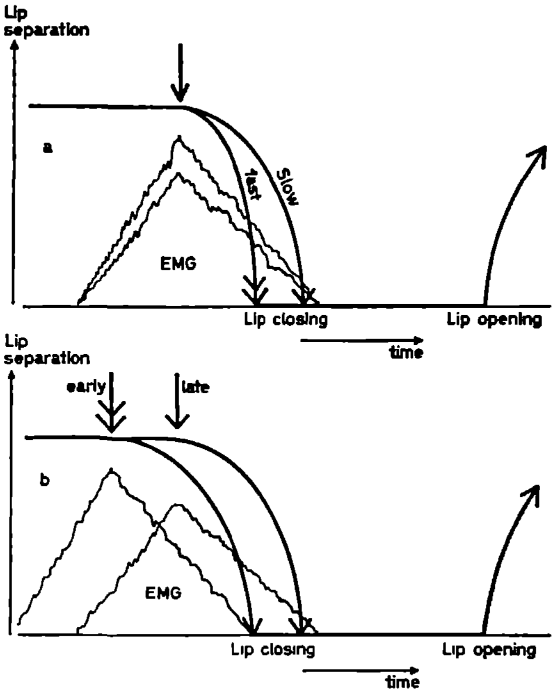


Fig. 2. Schematic representation of 2 hypotheses accounting for an advanced moment of lip closure with a stronger command (double arrow) compared with a weaker command (single arrow). *a* The 'moment' of innervation is the same, the speed of the closing movement is different. *b* The speed of the lip-closing movement is the same but the moment of innervation is shifted.

preciably. A combination of the 2 hypotheses might also be possible.

A comparison of the time intervals between muscle activity and lip closure in /p/ and /b/, stressed and unstressed /p/, /p/ after short and long vowels and /b/ before long and short vowels should reveal which of the hypotheses more nearly fits the observed changes in duration best.

An interval which was shorter with the stronger command than with the weaker command would be indicative of a faster movement. An equal interval for strong and weak commands would be indicative of equal speed of the closing movement.

With the opposition between voiceless and voiced consonants (/p/ vs. /b/) and between stressed and unstressed syllables we do not observe any differences in the interval between the moment of the electromyographic peak and the moment of lip closure (table IV). The difference in closure duration proved to be about 15–20 msec in these cases. We therefore suggest that the moment of innervation is advanced in time with greater effort.

With lip closure after short and long vowels a slight tendency exists towards longer intervals between muscle activity and lip closure in short vowels than in long vowels. This might indicate a slower closing movement after short vowels than after long ones. On the basis of a longer closure duration of /p/ after short vowels and a higher electromyographic peak we expected the opposite to happen. We conclude that with this opposition the hypothesis that a stronger command results in a faster movement can be rejected. A shift in innervation must therefore occur to explain differences in closure duration of following /p/. The observed tendency towards a longer duration of the closing movement after short vowels is not explained.

With lip opening of /b/ in long and short vowels less consistent data were obtained than in the lip-closure measurements. The mean interval between the moment of the peak of the electromyographic activity and the moment of lip opening seems to be the same in long and short vowels. Further measurements are needed to obtain better results. The present data seem to indicate a mechanism similar to that involved in the lip-closure movement.

2. Discussion of the Oppositions in Effort

The results of the measurements will be interpreted in terms of a simplified model, represented as a black box. Its input consists of a series of programme units corresponding with 'features' that are bundled in single phonemes. We will use the term 'feature command' for the input units, which we hypothesize to be invariant. The feature command 'lip closure' is therefore assumed to be the same for /p/, /b/ and /m/ and independent of context. The output is the acoustic speech signal. Measured differences in duration and muscle activity between these phonemes and between different contexts will have to be added to the black box. The rules governing the changes might be called 'implementation rules' [Lieberman, 1968].

In the introduction we mentioned 4 different cases in which we expected effort oppositions on introspective grounds. Some terms indicating these introspective oppositions were also mentioned. The 4 oppositions are presumably somewhat inhomogeneous with respect to the origin of the differences in effort. We will discuss the 4 oppositions more extensively in this chapter.

a) Voiceless Consonants vs. Voiced Consonants

In an earlier article we tried to construct a model for the voiced-voiceless distinction which would explain all differences between voiced and voiceless consonants by a different use of one articulator, viz. the pharynx constrictor muscle [Slis and Cohen, 1969a, b]. However, the model did not account for differences in closure duration and in electromyographic activity of the lips in bilabial plosives. It was therefore extended with an assumed 'cross talk' between the 'command lines' from the brain to the articulators [Slis, 1968, 1970]. With voiceless consonants the command to the pharynx musculature would 'leak' into other command lines, which would result in stronger commands to the other articulators as well (e.g. to the lips in /p/). The increased muscle activity with voiceless consonants might introspectively be experienced as a higher articulatory effort, thus giving rise to the term 'tense' for voiceless consonants. In the simplified model described we assume that cross talk is part of the implementation, and in fact would lead to an earlier initiation of the consonant.

b) Stressed vs. Unstressed

The origin of the differences between stressed and unstressed syllables or between emphatic and normal speech is probably different. It seems reasonable to assume that this originates at a higher level in the speech production chain than does the voiced-voiceless opposition, and it is consequently added to our black box. This is equivalent to the 'prosodic effort feature' of Lieberman [1968] or to the 'fluctuating effort level' of Oehman [1967]. The increased effort would result in an overall higher activation of all articulators involved [Fónagy, 1966]. In fact a higher electromyographic activity with stressed compared to unstressed syllables was found for the internal intercostal muscles [Draper, 1959; Fónagy, 1966], the larynx muscles [Fónagy, 1966; Oehman, 1967; Perkins, 1968], the velar muscles [Fritzel, 1968] and the lip muscles [Harris, 1968]. Again the change from effort to higher activation takes place according to implementation rules.

c) Closure After Short Vowels vs. Closure After Long Vowels

In a recent article on the phonetic feature of vowel length (Nooteboom and Slis, subm. for publ.] we hypothesized that this stronger closing command is part of the mechanism for making vowels short. Thus implementation of the feature of vowel length also implies that the closure for the following consonant is made strong. The terms 'scherp gesneden' (Dutch) and 'scharf geschnitten' (German), both meaning 'sharply cut', and 'fester Anschluß' (German), meaning close contact, for short vowels and the terms 'zwak gesneden' (Dutch) and 'weich geschnitten' (German), both meaning weakly cut, and 'loser Anschluß' (German), meaning loose contact, for long vowels indicate a difference in articulatory effort on the end of the vowel. Differences in effort with the closing movement after short and long vowels are experimentally shown to be present in German by FISCHER-JØRGENSEN [1969] by means of measurements on the airstream, intraoral pressure and duration, respectively, just before, during and after the short or long vowel.

d) Opening Movement of /b/ Before Long Vowels vs. /b/ Before Short Vowels

Whether differences in the opening movement in long and short Dutch vowels have to be attributed to the feature of vowel length as well or whether we have to hypothesize a separate feature is difficult to decide. For the time being we think it appropriate to attribute differences in both opening and closing movement between long and short vowels to the feature of vowel length; long vowels have a 'tense' opening movement and a 'weak' closing movement and short vowels have a 'lax' opening movement and a 'strong' closing movement.

Long vowels are often called 'tense' (in Dutsch: gespannen) and short vowels 'lax' (in Dutch: ongespannen). PERKELL [1969] observed on cineradiographic films a more extreme deviation of the articulators from a 'neutral position' during long vowels than during short. In English MACNEILAGE and SHOLES [1964] measured a higher electromyographic activity in long vowels than in the corresponding short vowels, viz. /i/, /æ/, /a/ and /u/ vs. /ɪ/, /ɛ/, /ʌ/ and /ʊ/. A possible stronger opening command may be part of the mechanism to give the articulators a more extreme position in long vowels than in short ones.

e) Survey of the Effort Oppositions

We hypothesized different causes in the 4 effort oppositions mentioned above, viz. (1) for voiceless vs. voiced consonants we assumed cross talk of the command to the pharynx musculature, which is part of the implementation; (2) for stressed and unstressed syllables we assumed a different input from a higher speech-programming level; (3) for the closing movement after short and long vowels and (4) for the opening of the mouth in long and short vowels we assumed that the differences in effort are dependent on the feature of vowel length. The implementation of the latter 2 oppositions seems to be of a completely different character since long vowels are supposed to have a strong opening movement, which is an activation, and a weak closing movement, which points to an inhibition. From the results of the measurements of the moments of closing and opening of the lips, we found, in spite of different origins of the effort oppositions, an advancement in time of the stronger command in all 4 oppositions. We therefore think it probable that the most peripheral part of the black box model assumed above contains an implementation rule for effort that is used in all effort oppositions discussed.

VI. Conclusion

A survey of the results obtained shows that speech sounds or speech movements that are intuitively sensed as being executed with more effort than other movements in the same context are advanced in time, leading to a longer closure duration with an advanced moment of lip closure, viz. an earlier moment of lip-closure of /p/ than of /b/, of /p/ in stressed than in unstressed syllables, and of /p/ after a short vowel than after a long vowel, or leading to a shorter lip-closure duration with advanced moment of lip opening, viz. of /b/ before long vowels compared to /b/ before short vowels. Electromyographic measurements showed that the muscle activity of movements that are executed with more effort is higher than of those that are sensed as 'weaker', with the exception of the opening movement of /b/ with long and short vowels. Here, however, a possibly wrong electrode placement makes interpretation difficult.

The interval between the moment of electromyographic activity and the subsequent opening or closing of the lips proved to be the same or longer with strong movements than with weak movements. We therefore concluded that the advancing of the moment of lip opening or closing was not caused by a faster movement with a higher activity, but by an advancement of the muscle innervation with the stronger movement. We wish to place these results in the framework of a model in which invariant feature commands are the input of a black box which implements lower speech functions.

The implementation rules account for differences found between individual speech sounds with respect to context, e.g. differences in duration of lip closure with labial plosives in different contexts. The results obtained suggest an implementation rule that advances stronger commands in time, thus causing differences in duration of speech sounds that differ in effort. The differences in effort may orginate from different feature commands (voiceless vs. voiced consonants or long vs. short vowels) and in a different prosodic input (stressed vs. unstressed or emphatic vs. normal speech).

Summary

Measurements of lip-closure duration and electromyographic activity of the lips were done on: (1) labial plosives that were articulated with degrees of effort which were introspectively sensed to be different, and (2) labial plosives adjoining vowels that were articulated with degrees of effort which were introspectively sensed to be different. The measurements of closure duration showed a lengthening of closure duration with increasing effort of the plosives, i.e. /p/ is longer than /b/ and /p/ within a stressed syllable is longer than /p/ in an unstressed syllable. This can be interpreted as an earlier moment of lip closure with more effort. The /p/ after short (more effort) vowels is longer than after long (less effort) vowels and /b/ before long (more effort) vowels is shorter than before short (less effort) vowels. This is in accordance with an earlier execution of the movement with more effort. The electromyographic measurements showed a higher activity with all lip-closing movements that were expected to be executed with higher effort. With lip opening the electromyographic measurements yielded less convincing results, although a tendency to higher activity with more effort could be observed. The results obtained could be interpreted as an implementation rule of the articulatory system stating that commands with more effort are advanced in time compared to commands with less effort.

Zusammenfassung

Die Artikulationsleistung und ihre Wechselbeziehungen zur Dauer und Elektromyographie

Dieser Artikel behandelt introspektiv perzipierte Unterschiede der Starke der Artikulation. Es wurden Messungen der Schließungsdauer der Lippen vorgenommen. Die Resultate zeigten, daß das kräftige /p/ länger ist als das weichere /b/, daß das /p/ in einer betonten Silbe langer ist als in einer unbetonten und daß das /p/ hinter einem kurzen (scharf geschnittenen) Vokal länger ist als hinter einem langen (weich geschnittenen) Vokal. Ein /b/ vor einem langen (gespannten) Vokal ist kürzer als vor einem kurzen (ungespannten) Vokal. Diese Ergebnisse können interpretiert werden als eine Verfrühung der Öffnungs- und Schließbewegung bei größerer Starke. Die elektromyographischen Messungen zeigten eine höhere Aktivität bei der Schließbewegung von 1. /p/ im Vergleich zu /b/, 2. von /p/ in einer betonten Silbe im Vergleich zu einem in einer unbetonten und 3. von /p/ hinter einem kurzen Vokal im Vergleich zu einem /p/ hinter einem langen Vokal. Weiterhin war eine Tendenz zu einer größeren Aktivität bei der Öffnungsbewegung von /b/ vor einem langen Vokal im Vergleich zu einem solchen vor einem kurzen Vokal zu beobachten. Diese Ergebnisse können in dem Sinne interpretiert werden, daß eine «implementation rule» (Implementregel) bestehen muß, die besagt, daß eine höhere artikulatorische Stärke eine verfrühte Innervation der Muskeln hervorruft.

Résumé

L'effort articulatoire et ses corrélats de durée et électromyographie

Nous avons mesuré la durée de la fermeture des lèvres et de l'activité électromyographique des muscles des lèvres, pendant (1) des occlusives labiales, articulées avec des degrés d'effort intuitivement perçus comme différents et (2) des occlusives labiales suivant ou précédant des voyelles et articulées avec des degrés d'effort intuitivement perçus comme différents. Les mesures effectuées montrent un prolongement de la durée de l'occlusion parallèle à l'augmentation de la force de l'articulation, c'est-à-dire que le /p/ a une plus grande durée que le /b/ et que le /p/ a une plus grande durée dans une syllabe accentuée que dans une syllabe nonaccentuée. On peut interpréter cela comme un avancement du moment de l'occlusion lorsque l'effort est plus grand. Après une voyelle brève le /p/ est plus long qu'après une voyelle longue. Le /b/ avant une voyelle longue est plus bref qu'avant une voyelle brève. L'influence de la force de l'articulation de la voyelle sur la durée des occlusives voisines peut s'expliquer également par un avancement de l'exécution en fonction d'un effort plus grand. Les mesures électromyographiques montrent une plus grande activité pour tous les mouvements des lèvres demandant un effort relativement grand. Les résultats obtenus peuvent s'interpréter comme application de la régle du système articulatoire qui dit que les efforts d'articulation les plus intenses donnent lieu à des commandements neuraux donnés plus tôt que ce n'est le cas pour les efforts moindres.

References

Cooper, F. S.: Instrumentation methods for research in Phonetics. Proc. 5th Int. Congr. Phon. Sci., pp. 142–171 (Karger, Basel 1964).

Draper, M. H.; Ladefoged, P., and Whitteridge, D.: Respiratory muscles in speech. J. Speech and Hearing Res. *2:* 16–27 (1959).

Fischer-Jørgensen, E.: Untersuchungen zum sogenannten festen und losen Anschluß. Kopenhagener Germanische Studien, Vol. 1, pp. 138–164 (Akademisk Forlag, Kopenhagen 1969).

Fónagy, I.: Electrophysiological and acoustic correlates of stress and stress perception. J. Speech and Hearing Res. *9:* 231–244 (1966).

Fritzel, B.: The velopharyngeal muscles in speech. Acta oto-laryng., suppl. 250 (Stockholm, 1969).

Fromkin, V. A. and Ladefoged, P.: Electromyography in speech research. Phonetica *15:* 217–242 (1966).

Harris, K. S.; Gay, T.; Sholes, G. N., and Lieberman, P.: Some stress effects on electromyographic measures of consonant articulations. Haskins SR *13/14*: 137–151 (1968).

LIEBERMAN, P.: On the physical basis of phonological features. Haskins SR *15/16:* 137–155 (1968).

LINDBLOM, B.: Spectrographic study of vowel reduction. J. Acoust. Soc. of Am. *35:* 1773–1781 (1963).

MACNEILAGE, P. and SHOLES, G. N.: An electromyographic study of the tongue during vowel production. J. Speech and Hearing Res. *7:* 209–232 (1964).

NOOTEBOOM, S. G. and SLIS, I. H.: Timing of Dutch stressed words. I.P.O. Ann. Progr. Rep. *4:* 52–58 (1969).

NOOTEBOOM, S. G. and SLIS, I. H.: The phonetic feature of vowel length. Lang. Speech (submitted for publication).

OEHMAN, S. E. G.; MARTENSSON, A.; LEANDERSSON, R., and PERSSON, A.: Crico-thyroid and vocalis muscle activity in the production of Swedish tonal accents: A pilot study. STL-QPSR *2–3:* 55/57 (1967).

OEHMAN, S. E. G.: Word and sentence intonation: a quantitative model. STL-QSR *2–3:* 20–54 (1967).

PERKELL, J. S.: Physiology of speech production: Results and implications of a quantitative cineradiographic study (M.I.T. Press, Cambridge, Mass. 1969).

PERKINS, W. H. and YANAGIHARA, N.: Parameters of voice production. I. Some mechanisms for the regulation of pitch. J. Speech and Hearing Res. *11:* 246–267 (1968).

SLIS, I. H.: Experiments on consonant duration related to the time structure of isolated words. I.P.O. Ann. Progr. Rep. *3:* 71–80 (1968).

SLIS, I. H. and COHEN, A.: On the complex regulating the voiced-voiceless distinction I and II. Lang. Speech *12:* 80–102, 137–155 (1969).

SLIS, I. H.: Articulatory measurements on voiced, voiceless and nasal consonants: A test of a model. Phonetica *21:* 193–210 (1970).

Author's address: I. H. SLIS, Institute for Perception Research, Insulindelaan 2, *Eindhoven* (Netherlands)

CHAPTER 6

INTERMEZZO
DISCUSSION OF THE "EFFORT" MODEL OF THE VOICED-VOICELESS DISTINCTION

6.1 INTRODUCTION

In this intermezzo, the model of the voiced-voiceless distinction which was developed in the first four articles, will be discussed.
In the second section it will be shown that no new developments can be reported with respect to the acoustic features.
In the next section attention will be paid to the mechanisms underlying the difference in volume of the pharynx cavity; two possibilities will be considered: (a) a contraction of the pharynx cavity during voiceless obstruents, and (b) an active expansion of the pharynx cavity during voiced obstruents.
The fourth section is devoted to new data on a greater speed of the articulators during the closing movement of voiceless obstruents compared to voiced ones. These data corroborate the model with respect to the view that the difference in articulatory effort is an important aspect of the voiced-voiceless distinction.
In the fifth section aspiration will be discussed. Although phonologically not relevant for Dutch, aspiration forms an element in the discussion of the voiced-voiceless discussion in the international literature.
In the final section of this intermezzo the model of the voiced-voiceless distinction will be revised on the basis of electromyographic evidence published after the fourth article appeared. This evidence concerns measurements showing an active glottal opening gesture of the glottis during voiceless obstruents.

6.2 THE VOICED VOICELESS DISTINCTION: ACOUSTIC FEATURES

In Slis and Cohen (1969 a (first article in this volume)) an example of our segmental time domain approach of the voiced-voiceless distinction was given. All known acoustic aspects accompanying the voiced-voiceless distinction that were found in the current literature on different languages were discussed. We added data on Dutch that were measured in spectrograms and oscillograms, and that were obtained by perceptual experiments with the IPOVOX-I and II. The description of these aspects were, as far as possible, given in terms of acoustic segments. In Slis 1970 (3rd article in this volume) these aspects were summarized in eight points, viz.:

1) duration of the consonant,
2) duration of the preceding vowel,
3) duration and spectral extensiveness of the vowel formant transitions,
4) presence or absence of vocal vibration during the consonant,
5) duration of the noise burst during plosives,
6) sound pressure of noise burst in plosives and friction noise in fricatives,
7) sound pressure of the adjoining vowels,

8) peak value of the fundamental frequency of the surrounding vowels and the contour of the fundamental frequency during the following vowel.

At the same time a similar list of acoustic and articulatory differences between voiced and voiceless plosives in French and Danish was published by Fisher-Jørgensen (1969). The list for French resembled the list given above. In addition to the points mentioned, a faster intensity rise in the following vowel after voiceless than after voiced plosives was observed. This finding was confirmed by Debrock (1977) for Dutch and French obstruents.
A second additional acoustic difference observed by Fisher-Jørgensen concerns the duration of the following vowel. She defined vowel duration on an acoustic basis; in a succession of a voiceless plosive and a vowel this boils down to the time interval between the voice onset and the sudden decrease of the amplitude caused by the closing gesture at the end of the vowel. The time interval between the moment of oral opening and the moment of voice onset (VOT) is regarded to be part of the consonant, and called "open interval". In terms of this volume, however, vowel duration was defined on an articulatory basis, viz. the time interval between the oral opening and oral closing. This amounts to the sum of the open interval and the vowel duration as defined by Fischer-Jørgensen. She observed that the acoustic vowel duration is shorter after voiceless than after voiced plosives. The difference in vowel duration is (partly) compensated by a difference in the open interval.

The non-segmental frequency domain approach of the Haskins investigators is illustrated by a similar list of 16 points (Lisker 1978). The 16 cues in the example were actually measured in the word pair "rapid-rabid". For the sake of comparison, the order in Lisker's list was changed, and his cues were regrouped and renumbered; the second number refers to the corresponding point in the Slis (1970) list given above:

1)-1 duration of consonantal closure,
2)-2 [ae]-duration (preceding vowel),
3)-3 a. F1-offset frequency,
4)- b. F1-closing transition duration,
5)- c. F1-cutback before closure,
6)- d. F1-onset frequency after closure,
7)- e. F1-opening transition duration,
8)- f. F1-cutback following closure,
9)-4 a. presence/absence of low frequency buzz during the closure interval
10)- b. timing of voice offset before closure,
11)- c. decay time of glottal signal preceding closure,
12)-5 VOT-delay after closure,
13)-6 intensity of burst following closure,
14)-7 amplitude of [I] relative to [ae],
15)-8 a. F0-contour before closure,
16)- b. F0-contour after closure.

The aspects numbered 1, 2, 9, 13, and 14 are based on a segmental description, the other aspects are defined in the time continuum with respect to moments of closing or opening of the vocal tract, or spectral

values at a fixed point in time.
A closer comparison of Lisker's and our list of acoustical aspects accompanying the voiced-voiceless distinction shows that both lists contain essentially the same information; Lisker went more into detail, especially where F1-transitions are concerned. With respect to the presence or absence of voice during the closure interval, Lisker's list contains two cues regarding voice-offset (10 & 11). These cues are not relevant for intervocalic or final stops in Dutch, since final stops are always voiceless and intervocalic voiced stops have uninterrupted voice activity in Dutch. Consequently no differences between voiced and voiceless stops are cued by differences in timing of voice offset and the decay time of the glottal signal before closure. However, if the post-vocalic stop is the first consonant of a cluster in which assimilation may occur, the timing of voice offset can be used as a criterion to define "voicing" of a post-vocal consonant.

6.3 ADDITIONAL REMARKS ON PHARYNGEAL VOLUME CONTROL

In most models of the voiced-voiceless distinction, the higher intra-oral pressure during voiceless obstruents is explained by a difference in volume of the pharyngeal cavity. In the model presented in this volume it was assumed that the pharynx cavity is contracted during voiceless obstruents, and not active during voiced ones. During voiced obstruents a passive expansion may take place because of the elasticity of the pharyngeal walls; the pharynx is, as it were, blown up. EMG measurements on the pharynx constrictor muscle corroborate this hypothesis. This muscle shows less activity during voiced plosives than during voiceless ones (Minifie et al. 1974).

At the basis of an alternative explanation for the intra-oral pressure differences between voiced and voiceless obstruents, lies the presupposition that an active expansion of the supra-glottal cavity takes place during voiced consonants, instead of (or combined with) a contraction during voiceless ones. This would result in similar pressure differences. Muller (1983) presented calculations made with a model of Rothenberg, indicating that expansion during voiced obstruents is necessary to create a supra-glottal volume that is large enough to receive a glottal airflow needed to sustain vocal fold vibration. Muller (1983) calculates that if the supra-glottal cavity is not allowed to expand passively because of an increasing intra-oral pressure, the time constant for the decay of glottal airflow will be roughly 55 to 120 ms. The requisite airflow for sustaining voicing will be roughly half of this time constant, viz 30 to 60 ms. In case the walls of the pharynx are not yielding in [+ tense] consonants, voicing will cease within 50 ms in most cases. Voicing will be resumed at the opening of the vocal tract. Comparable estimates were made by Keating (1984) with a model which was based on the Rothenberg approach. Abramson (1976) reports even shorter durations which were also obtained with the Rothenberg model, viz. 4 ms with no expansion of the vocal tract, 20-30 ms with passive expansion and 80-90 ms with active expansion. Longer voicing should be due to nasal leakage. These data are obtained with a relatively simple model in which the vocal cords are regarded as "one mass" and in which static transglottal pressures are assumed; it is assumed that vocal vibration is interrupted if the transglottal pressure falls below 2 cm water pressure (Catford 1977). Recent investigations do not bear out these assumptions (Cranen 1983, Cranen and Boves 1984). These estimates are ,therefore, to be regarded with reservation.

However, no suggestions were given with respect to the mechanisms that could perform active enlargement.
The muscles in the back and lateral walls of the pharynx are not attached to any points outside the vocal tract that are far enough backward and/or lateral to accomplish an expansion. Active enlargement in that direction seems therefore improbable.
Another possibility to enlarge the supra-glottal cavity lies in a forward movement of the tongue or in a lowering of the jaw (Perkell 1969). A forward movement of the tongue pulls a chain that consists of the hyoid bone and the superior horns of the thyroid cartilage. Pulling this chain leads to a forward tilting of the thyroid cartilage, which effects a stretching of the vocal cords. In that case the pitch in vowels adjacent to voiced obstruents should rise instead of fall (Honda 1983). Halle and Stevens (1969) also doubt this possibility. They say in this context that "... a different mechanism may well be at work in these consonants."
Evidence is presented with respect to a different jaw position by Fujimura and Miller (1979) who observed that the jaw movement in the closing gesture of final /t/ is faster and larger than of final /d/. This has to result in a higher jaw position in /t/ than /d/, and consequently a smaller oral cavity in /t/ than in /d/. Keating (1983:403) measured that " The jaw positions from the highest to the lowest are /ztsfvjpnblmgwrk/"; in agreement with Fujimura and Miller we see that in /f/ and /p/ the jaw was higher than in /v/ and /b/. On the other hand /s/ and /k/ yielded lower jaw positions than /z/ and /g/. Keating referred to data from Lindblom who suggested that jaw position is lower with increasing sonority. These data on jaw positions are too accidental to base a firm conclusion on. However they seem to indicate that a lower jaw position in voiced obstruents might be one of the causes of a difference in supra-glottal cavity size.
A third possibility to create supra-glottal enlargement lies in lowering the glottis. Lindqvist et al. (1973) showed laryngeal lowering in voiceless and voiced stops. In voiced stops these laryngeal movements resulted in lower larynx positions than in voiceless ones. Ewan and Krones (1974) measured larynx height in English, Thai, French and Hindi. They observed no consistent lowering of the larynx in consonants, but that " Consonants are generally higher or lower in larynx position agreeing in relative position of the vocalic environment" and that voiceless stops generally have a higher larynx position than corresponding voiced stops. The difference in height between voiceless and voiced stops was generally between 1 and 2 mm, which can hardly cause a substantial difference in supra-glottal volume.

These data on alternative explanations for the different volumes of the supra-glottal cavity in voiced and voiceless obstruents are not firm enough to justify a change in our original hypothesis that the difference is caused by an active contraction of the pharynx in voiceless obstruents and a passive expansion during voiced ones.

However, if we assume that an active pharynx expansion is worked by a forward tongue movement or by a lowering of the jaw and/or the larynx, we have to keep in mind that this additional movement cannot be explained by a difference in effort. In that case the voiced-voiceless opposition is complicated by an additional articulatory feature, "pharyngeal expansion".

6.4 ADDITIONAL REMARKS ON ARTICULATORY SPEED

Recent data in the literature showed that the speed of the articulatory movements is greater in voiceless than in voiced obstruents. Fujimura and Miller (1979) showed that the upward (closing) movement of the jaw is faster in final /t/ than in final /d/. Kohler (1980) reported that the velar closing gesture in /t/ is faster than in /d/. The observation of Debrock (1977) that "... in 53 out of 60 times, the decay time of the vowel which precedes a fortis is less long than in the case of the corresponding lenis ..." can be explained by a faster closing movement in voiceless (fortis) obstruents, since the rate of amplitude change can be regarded as a correlate of articulatory speed.
Kohler (1980) observed that the increase of the speed of the closing movement at the onset of fortis obstruents was accompanied by a decrease of the speed of the opening movement of the subsequent vowel. This evidence makes it improbable that the faster amplitude rise after voiceless obstruents (observed by Debrock, 1977) is due to a faster opening movement. It seems more likely that the difference in amplitude rise is the result of a delayed voice onset after voiceless obstruents compared with voiced ones. If the VOT is delayed, the oral aperture will be larger at the VOT and the sound emission of the subsequent vowel will be less damped; as a result the initial amplitude will be larger at the vowel onset. Consequently the rise time of the amplitude of the subsequent vowel will decrease with a delay in VOT.

In an article on the fortis-lenis opposition in German, Kohler (1977:45) stated that "... highly significant negative correlations of vowel and stop durations can be interpreted by relating them to the underlying principle of force or speed of articulatory movement ". Measurements of the EMG-activity of the palatoglossus muscle in /k/ and /g/ showed that the increase in activity in the /g/ occurred much later than in the /k/ "... the mean delay in the utterances being t = 63 ms..." (Kohler 1977:50). The maximum activity was reached at the beginning of the oral closure. These data indicated that presumably both explanations hold: more effort leads to faster movements and to an advancement of the EMG-activity.

6.5 ADDITIONAL REMARKS ON ASPIRATION

In several articles, Fischer-Jørgensen reported on differences along the "voicing dimension" in various languages. She showed that both /b,d,g/ and /p,t,k/ were realized as voiceless plosives in several languages among which is Danish (Fischer-Jørgensen 1969, 1979, 1980). Although the maximal degree of glottal opening was larger during /p,t,k/ than during /b,d,g/, /b,d,g/ were found to be produced with an open glottis as well. The main difference was formed by a later moment of glottal opening in /p,t,k/. Maximal glottal opening in /p,t,k/ was reached at the moment of oral release. In /b,d,g/ the glottis was already closed again at the moment of oral opening. The difference in glottal timing led to voiceless aspirated /p,t,k/ and to voiceless non-aspirated /b,d,g/. Similar results were obtained by Petursson (1976) in Icelandic. Since both groups of plosives are voiceless he called them /ph,th,kh/ and /p,t,k/ instead of /p,t,k/ and /b,d,g/.

On the basis of her results Fischer-Jørgensen (1969, 1980) argued that the features "effort" and "aspiration" can be used independently. In Danish [+ aspiration] is combined with [- tense], while on the other hand in e.g. English [+ aspiration] is coupled with [+ tense].

In summary, the timing of the glottal opening and closing movement and that of the oral opening and closing movement can, evidently, be controlled separately, thus leading to various VOT's. Differences in VOT between two voiceless (open glottis) consonant classes can be described by an articulatory feature "aspiration", which is in fact a difference in glottal timing. Differences in VOT between a voiced (closed glottis) and a voiceless (open glottis) consonant can be described by an articulatory feature "voice", which is in fact a difference in glottal control. This will be discussed in more detail in the next section.

Differences in VOT also occur between comparable consonant classes in different languages. These differences are sometimes small (Lisker and Abramson 1964). In these cases one can hardly speak of a difference caused by an articulatory feature, although timing differences occur. These points will be discussed in the epilogue.

6.6 ADDITIONAL REMARKS ON LARYNGEAL CONTROL REVISION OF THE MODEL OF THE VOICED-VOICELESS DISTINCTION

At the time the article which proposed the model described in this volume was published, other possibilities were proposed to explain the interruption of voicing or the delay of the VOT in voiceless obstruents. Lisker and Abramson (1971) made two different suggestions. The first stated that the differences in VOT between voiced and voiceless obstruents were a result "... of varying the time of arrival of neural motor signals to the appropriate laryngeal muscles to close the glottis ..." (1971:770). As an equally probable explanation they suggested that it might" ... involve changes in the balance of forces exerted by the various muscles in and upon the larynx." (1971:770).

Shortly after the appearance of the four articles in which an effort model for the voiced-voiceless distinction was proposed, evidence became available that laryngeal control was different in voiced and voiceless obstruents. Therefore, an explanation in terms of more or less effort with an otherwise identical innervation program for the muscles involved was no longer tenable.

Hirose and Gay (1972) reported EMG measurements on the intrinsic laryngeal muscles during speech, performed with hooked-wire needle electrodes. They observed that the posterior cricoarytenoid muscle (PCA) showed a consistent increase of activity for voiceless consonant production; the increase was absent for voiced consonants. Simultaneous measurements on the internal arytenoid muscle (INT) indicated that there was reciprocal activity between PCA and INT. They concluded that the

role of the PCA can be considered to be that of the abductor of the vocal folds (opens the glottis). The INT on the other hand acts as the adductor of the vocal folds (closes the glottis).
Collier et al. (1979) produced EMG data of intrinsic laryngeal muscles during speech in Dutch. They showed that before a voiceless obstruent the INT activity started to decrease; this was not observed (or observed to a lesser degree) before a voiced obstruent. The lateral cricoarytenoid muscle (LCA) showed no distinct differences in activity. The activity of the vocalis (VOC) tended to be somewhat higher after a voiceless consonant. Unfortunately, their measurements of the PCA deteriorated and could therefore not be processed. Nevertheless, their data seem to confirm those of Hirose and Gay (1972).
Sawashima (1979) gave a review of the data obtained from various languages all showing a clear reciprocal pattern of activity between tha PCA and the INT muscles for the voiced-voiceless distinction, viz. an increase in PCA activity and a decrease in INT activity for voiceless sounds, and the reverse for voiced ones.
The model presented in this volume (which is summarized in the introduction section 1.4) was critisized on the basis of this new evidence (Sawashima et al. 1970). These laryngeal data therefore make a further revision of the model necessary. A different laryngeal innervation pattern for voiced and voiceless consonants has to be included. This difference in laryngeal action provides an alternative explanation for some of the articulatory and acoustic events that accompany the voiced-voiceless distinction, such as interruption of voicing during voiceless obstruents, a higher intra-oral pressure and a stronger friction noise in voiceless obstruents as compared to voiced ones.

In addition to the programmed difference in glottal aperture by means of the intrinsic laryngeal muscles, the larynx as a whole has a different position depending on the voice character of the produced consonant. The properties of the vocal cords are dependent on the larynx position; this has consequences for the fundamental frequency and the spectral composition of the larynx signal. The exact relation between the condition of the vocal folds, the position of the larynx and the mechanisms acting upon this position is still a matter of hypotheses (Erikson et al. 1982) which does not seem relevant enough to be discussed here. We will confine ourselves here to the observation that some of the extrinsic mechanisms that act on the larynx position have consequences for the frequency of vocal fold vibration and the amplitude of the signal.
Moreover, we are confronted with a number of other events which cannot be covered by a difference in laryngeal activation, viz. a longer duration of oral closure, higher EMG activity in various supraglottal muscles and a smaller pharynx cavity in voiceless obstruents. The latter facts can be explained by a mechanism as proposed in the model described above (Slis 1971 a). In order to update this model, a separately programmed opening and closing gesture of the glottis for voiceless obstruents is added on top of the increased effort, which was the only cause that was held responsible for the voiced-voiceless distinction in the model proposed originally. Events that are explained by more effort as well as opening of the glottis, such as high intra-oral pressure and

high pitch for voiceless obstruents, do not lead to contradictory results in both cases.

In the light of the fact that at least two (or four if pharyngeal expansion and aspiration are included) different mechanisms play a role, it is not surprising that the opposition between the group /p,t,k,s,f,x/ and the group /b,d,z,v/ give rise to two different phonological labels. The two mechanisms responsible for the voiced-voiceless distinction are in agreement with the phonological notions tense-lax regarding articulatory effort, and voiceless-voiced regarding laryngeal control. Whether these two mechanisms must lead to two different, independent phonological features will be dealt with in the light of assimilation of voice.

CHAPTER 7

ASSIMILATION OF VOICE IN RELATION TO VOICE QUALITY

I. H. Slis

Reprint from:

SOUND STRUCTURES
Edited by M. van den Broecke, V. van Heuven and W. Zonneveld
Dordrecht: Foris Publ., 1983, 245-257

1. INTRODUCTION

The present study on the influence of voice quality on assimilation of voice is part of a larger series of investigations dealing with various factors influencing assimilation in Dutch.[1] The factors examined so far are speaking rate, stress, phonological composition of the clusters, experimental condition, sex and dialectal origin of the speakers (van Erp, Halfens, Hoogeveen & Zink, 1981; Slis, 1981a,b,c,d). In one of these studies it was found that male and female speakers showed significant differences in assimilation of voice (Slis, 1981c). We proposed two hypotheses to explain this difference:

a. A sociolinguistic origin: assimilation of voice is learned behaviour that is conditioned by sociolinguistic norms.

b. An organic origin: differences between the vocal organs of men and women yield differences in assimilation of voice in male and female speakers.

The present research aims at demonstrating that the second hypothesis is correct. Male and female larynges differ with respect to size, stiffness and mass of the vocal cords. These organic-mechanical differences cause a difference in the boundary conditions under which vocal cord vibration occurs. On the same grounds we may expect similar phenomena in case the vocal organs differ with respect to other organic causes than those due to the sex of the speakers insofar as they lead also to differences in stiffness and/or mass of the vocal cords. This is for instance the case in speakers with a "poor" voice (due to organic aberrations) as compared to speakers with a "good" voice. The boundary conditions for phonation may also vary if a different use is made of the vocal cords, such as speaking on different pitches, which is mainly caused by a variation of tension and consequently of stiffness of the vocal cords.

If the free emission of air from the mouth is hindered by an oral constriction, as during obstruent articulation, the air pressure behind the obstruction will rise and the difference between intra-oral and subglottal pressure will decrease. As a consequence the flow rate of the transglottal airstream will decrease as well. In "poor" voices, with relatively unfavourable boundary conditions for phonation, vocal vibration will stop at a larger pressure difference across the glottis than in "good" voices, with optimal boundary conditions. Therefore we expect earlier and more frequent interruption of phonation during consonants in poor than in good voices. In a succession of a voiceless and a voiced obstruent therefore, more and longer voiceless intervals will occur in poor than in good voices. This results in more completely voiceless clusters (progressive assimilation) in poor voices.

Likewise we expect more progressive assimilation if the vocal cords are tense (high stiffness), as is the case in high pitched speech, relative to a condition in which the vocal cords are relaxed (low stiffness), as in low pitched or normally intonated speech. Halle & Stevens (1971) calculated the boundary conditions for phonation as a function of stiffness and glottal width. They argue that under all conditions vocal vibration is maintained longer with a lower pressure drop across the glottis with slack than with stiff vocal cords.

In the present experiment we compare speech of subjects with good and with poor voices. The group with poor voices consisted of two subgroups, viz. a group of subjects before and a group after some training sessions by a speech therapist. The second variable in our experiment is vocal pitch; we compared normally intonated speech and speech on a monotone on three different pitches. Besides a determination of degree and direction of assimilation of voice, we will try to find spectral differences between the speech of the groups that participated, and between speech obtained under different experimental conditions.

2. SUBJECTS

On the basis of the screening by a speech therapist we selected three classes of male subjects, viz. "good" voices, "untrained poor" voices and "poor" voices after a few (4-7) training sessions. Voices were defined "good" if they scored positive on four criteria, viz.:

1. A "large" voice: the voice output measured at a distance of 10 cm in front of the mouth ranges between over 104 dB(A) (Pahn, 1976) and below 50 dB(A).
2. A "normal" Pitch: one or more of the following three procedures were used to test "speaking pitch":
   a) Normal pitch lies a third above the lowest "good sounding" pitch a subject can produce.
   b) Normal pitch lies a fifth above the lowest pitch possible, regardless of the perceived voice quality.

c) A third test applied is Gutzmann's pressure test (e.g. Luchsinger & Arnold, 1965); if the pitch drops when the larynx is pressed backwards during normal phonation, speaking pitch may be considered to be too high.

3. A "normal pitch range": we decided that speakers with a good voice ought to be able to produce a range of at least two octaves.
4. A subjective diagnosis of "good voice": this diagnosis, provided by a speech therapist, was based on other, subjective, criteria, such as the amount of breathiness, hoarseness and resonance present in the voice, or on complaints of the subject about his voice. The three criteria above (1, 2 & 3) also contributed to his diagnosis.

Voices were classified as "poor" if the subjective diagnosis and at least one of the other three criteria were negative. The poor voices were classified in two subcategories, viz.:

a) a group that received no training or treatment after screening (untrained poor voices),
b) a group that attended at least four training sessions organized to improve their voices (trained poor voices).

We made audio and electrolaryngograph recordings of 42 male subjects, comprising 11 good voices, 12 untrained poor voices, 11 trained poor voices and 8 voices which could not be classified by our criteria; the latter 8 voices were not used in the comparisons between different voices, but they were added in the comparison between different pitches.

3. ACOUSTIC ANALYSIS

Besides assessing voice capacity by means of diagnostic tests, we performed acoustic analyses to define the speech material used. To this end a long time average spectrum (LTAS: Boves & Cranen, 1982) was made of the first 10 seconds of the recordings in each of the four speaking conditions (viz.: intonated, low-, medium- and high monotone, see below), both of the microphone and the electrolaryngograph signal, of all 42 subjects. For this we used a Bruel & Kjaer filter bank with bandwidths of about a critical band (van Rossum, 1981). Data reduction of the 20 filter outputs was obtained by applying a principal component analysis (IBM SSP, 1970). Only the first two components of both the microphone and the larynx signal proved to be interpretable and were retained for further analysis:

1. With respect to the microphone signal, the first dimension can be interpreted as the spectral slope above 1600 Hz and with respect to the larynx signal as the slope between 800 and 4000 Hz; above 4000 Hz noise plays a major role in the larynx signal. This first dimension explains about

30% of the variance of the microphone signal and about 50% of the larynx signal.

2. The second dimension seems to represent spectral intensity in the 650 and 3150 Hz range in the microphone signal (i.e. the F2-F3 region) as well as in the larynx signal. This second dimension explains an additional 20% of the variance of the microphone signal and 25% of the larynx signal.

The component scores on the two dimensions were used for a multivariate discriminant analysis (Overall & Klett, 1972). The weighting function that separated best between the groups with different voice qualities was used for a classification program (Overall & Klett, 1972). We compared the percentages of correct classification of category pairs, viz. good vs. untrained poor voices, good vs. trained poor voices and untrained vs. trained poor voices, with percentages of correct classification of pairs of (about equal size) random selected groups. This was done for both microphone and larynx signals for the four intonation conditions separately. We compared category pairs to be able to detect which categories can be separated on the basis of their spectra and which cannot. Classification of two randomly selected groups will presumably be better than 1:2 on the basis of differences in individual spectra. Therefore we used a paired t-test to decide whether classification of the experimental groups was significantly different from that of random groups instead of the classification of the category groups directly on a 50% chance basis.

4. RESULTS OF THE ACOUSTIC ANALYSIS

At the outset of the acoustic measurements we hoped to find substantial spectral differences between good and poor voices, since quality differences were clearly audible. The differences we found, however, are much smaller than expected. Still, some of the results seem worth mentioning.

A direct comparison between the LTA-spectra of the microphone signal of good and of both trained and untrained poor voices, shows somewhat lower spectral energy in the frequency interval between 600 and 1250 Hz in poor voices than in good voices for all intonation conditions. In figure 1 this is shown for the data of all four conditions added. No clear differences can be observed between the LTA-spectra of the larynx signal.

An inspection of the weighting functions of the multivariate discriminant analysis shows that, for the microphone signal, the second dimension of the principal component analysis is more important than the first; this is the spectral region between 650 and 3150 Hz. With the larynx signal we find the opposite, viz. that the first dimension, representing a spectral slope, is more important than the second, representing the F2-F3 region.

Classification based on the first two dimensions of the microphone

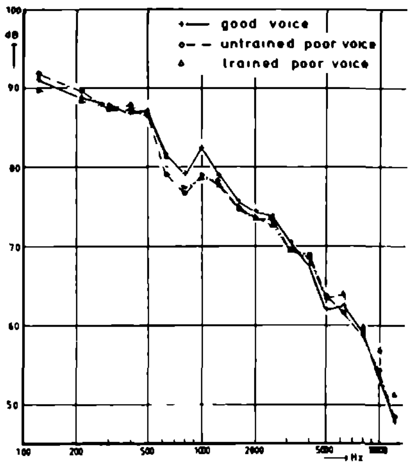


Figure 1. LTA-spectrum of the microphone signal of speakers with a good voice (N=11), an untrained poor voice (N=12) and a trained poor voice (N=11) in terms of the output levels of 19 Bruel & Kjaer (critical) band filters, which are normalized with respect to the overall speech level (100 dB).

signal separates good and poor voices (both trained and untrained) significantly better than the two random groups (cf. Table I on p. 250); trained subjects cannot be separated from untrained ones. The same procedure with the larynx signal only yields a separation significantly better than chance between good and untrained poor voices.

From these measurements we conclude that the groups of poor and good voices can be separated on the basis of acoustic measurements. The main difference seems to lie in the frequency region between 600 and 1250 Hz. A few training sessions by a speech therapist do not lead to a significant change in the voice spectrum.

Classification is better on the basis of the microphone signal than on the basis of the electrolaryngograph signal despite the additional "noise" due to the influence of the varying supraglottal cavities during speech.

5. SPEECH MATERIAL AND EXPERIMENTAL METHOD

The speech material consisted of 12 sentences. In each sentence a prestressed two-consonant cluster was present in which assimilation of voice could occur across word boundaries, viz. /pd,td,kd,pb,tb,kb,fd,sd,χd,fb,sb, χb/. First the subjects were asked to read the sentences aloud (intonated version). After that we presented the subjects a buzz (saw tooth function) with a fundamental frequency of 110 Hz via head-phones. They were

Table I. Percentage of correctly classified cases by discriminant analysis on the basis of the first two dimensions of a principal components analysis of (a) the microphone signal and (b) the electrolaryngograph signal.

(a) microphone signal	good vs. poor untrained	good vs. poor trained	poor untrained vs. poor trained	random groups
normal intonated speech	82.6	77.3	60.9	51.9
low pitched monotone	82.6	81.8	60.9	55.5
medium pitched monotone	82.6	77.3	52.2	62.3
high pitched monotone	73.9	81.8	60.9	49.2
mean classification	80.4 *	79.6*	58.7	54.7
(b) electrolaryngograph signal				
normal intonated speech	69.8	54.5	60.9	58.4
low pitched monotone	78.3	54.5	56.2	58.9
medium pitched monotone	73.9	54.5	52.2	51.8
high pitched monotone	82.6	72.7	60.9	67.0
mean classification	76.1 *	59.1	57.6	59.0

* degree of significance of the difference between classification of the experimental groups and classification of the random groups; $p < .01$ (paired t-test).

asked to read the same 12 sentences on a monotone at the pitch they heard (low version). This was repeated with 130 and 150 Hz (middle and high version, respectively). Two of the subjects could not produce the 150 Hz monotone; they were asked to do an additional run at a pitch of 100 Hz, which was then considered to be their low version. This way we obtained an intonated version and speech at three different pitches from all our subjects.

The choice of the pitches of 110, 130 and 150 Hz was based on a pilot experiment in which subjects were offered pitches between 100 and 170 Hz. None of the speakers in the pilot experiment reported problems with the pitches of 110, 130 and 150 Hz. All experienced 150 Hz to be high.

We compared the frequency and the direction of assimilation in the four modes of intonation for all 42 subjects. With respect to the influence on assimilation of good vs. poor voices we compared the 11 good, 12 untrained poor and the 11 trained poor voices with each other.

Assimilation of voice was scored with the help of oscillograms in three categories (Slis, 1981a). The criterion for deciding whether the first consonant of the cluster was voiced or not was derived from measurements on intervocalic voiced plosives in which we observed a continuation of phonation (=voice tail) after the moment of oral closure of about 10-20 ms (s.d. 10 ms) (Slis, 1970). Therefore we defined the first consonant to be voiceless if the interval between oral closure (viz. the moment of sudden amplitude drop in the oscillogram of the microphone signal) and cessation of phonation (as observed in the oscillogram of the electrolaryngograph signal) was less than 50 ms (voice tail < 50 ms).

After a voiceless first consonant we defined the second consonant to be voiced if the voice onset preceded the consonant release (negative VOT). In several cases the voicing was interrupted a second time; we hypothesized that this was the result of equalization of the pressure above and below the glottis and that this interval was meant to be voiced. The second consonant was defined to be voiceless if the VOT was zero or positive.

If the first consonant was voiced, we observed in the majority of the clusters a continuation of voicing during the complete cluster interval. Evidently, the second consonant was voiced in these cases. However, in some instances we found an interruption of voicing after voice tails longer than 50 ms. On the same grounds as with a second interruption of voicing above, we assumed that this was due to aerodynamic conditions, and that the second consonant was to be regarded as voiced.

On this basis we arrived at the following three assimilation categories (depicted in fig. 2, p. 252):

1. No assimilation: voice tail shorter than 50 ms and negative VOT
2. Regressive assimilation: voice tail longer than 50 ms and zero or negative VOT; this includes continuation of voicing.
3. Progressive assimilation: voice tail shorter than 50 ms and zero or positive VOT

Besides, we measured the consonant durations between the moments of oral closing and opening from oscillograms.

6. RESULTS AND DISCUSSION OF THE ASSIMILATION MEASUREMENTS

In table II (p. 252) we observe that the frequency of assimilation increases (no assimilation decreases) going from normally intonated, via low and middle to high pitched monotonous speech. This sequence corresponds with the order in which the sentences were spoken by the subjects. Additional measurements of the intersentence pause durations show that speech tempo increases with every repetition; the fraction "pause duration" (viz. interruptions of voicing longer than 300 ms) of the first 10 seconds of each series decreases from 42% in intonated speech, via 32% and 27% to 25% in high pitched monotonous speech. This increase in tempo was to be expected since our subjects became more familiar with the material each time they spoke a sentence.

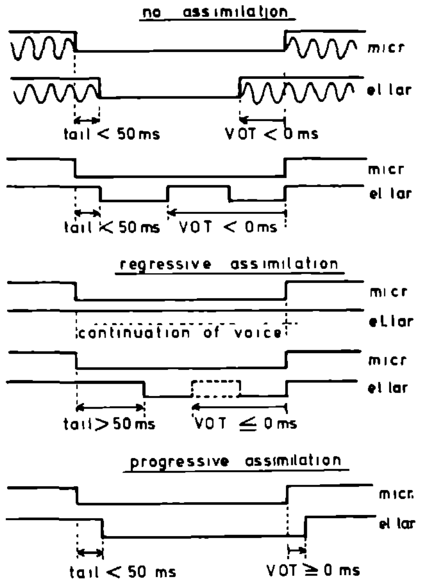


Figure 2. Schematic representation of the definitions of "no assimilation" "regressive assimilation" and "progressive assimilation".

Table II. Frequency of assimilation of voice in normally intonated speech and monotonous speech on a low, medium and high pitch of 42 male speakers. Parenthesized frequencies based on chance.

	speech with	monotonous speech on a:		
	intonation	low pitch	medium pitch	high pitch
No assimilation	108 (78.4)	80 (78.5)	70 (78.5)	56 (78.5)
Regressive ass.	347 (320.3)	323 (320.9)	326 (320.9)	287 (320.9)
Progressive ass.	48 (104.3)	101 (104.6)	108 (104.6)	161 (104.6)

$\chi^2 = 85.51$, df=6, $p < .001$.

Clusters with no assimilation are significantly longer than assimilated clusters (Table III). It seems that speakers use assimilation as a means of time compression. Articulation rate, however, seems to be equal under nearly all conditions for the classes of regressive and progressive assimilation, since no significant differences in cluster duration are found. We observe one exception, namely a significantly longer duration in progressive assimilation on a low pitch (Table III), for which we can offer no explanation.

Table III. Mean cluster durations (in ms) in speech with normal intonation and on a monotone on low, medium and high pitch. In parentheses: standard deviations (in ms).

	speech with intonation	monotonous speech on a: low pitch	medium pitch	high pitch
No assimilation	230 (65)*	218 (57)	207 (54)	202 (52)
Regressive ass.	156 (31)	158 (30)	153 (31)	154 (34)
Progressive ass.	151 (52)	179 (45)**	164 (34)	159 (35)

* duration with intonation significantly longer than on a monotone on a medium and a high pitch; $p < .02$ (t-test).
** duration significantly longer than all three other durations; $p < .01$ (t-test).

From former experiments, in which we measured the closure duration of successive stops in clusters in three-syllable nonsense words by means of mechanical contacts, we concluded that plosive overlap (i.e. unreleased first plosive) occurred (Slis, 1972). On the other hand, we observe in some of the tracings in the present experiment that a pause is introduced between the two consonants of a cluster in a few cases. In general we cannot decide on the basis of the voiceless intervals in the oscillograms whether overlap of the consonants takes place or separation after an unreleased plosive is present. In clusters with no assimilation we see that an increase in tempo (as observed from an increase in frequency of assimilation which accompanies the increase in pitch in the present experiment) results in shortening or omission of pauses, or by an increase of the overlap. In other words, the observed increases of the frequency of assimilation (decrease of no assimilation, Table III) with pitch is attributed to an increase in speech tempo, which is obtained by decreasing the pauses and by increasing overlap between successive consonants.

In our present experiment we find an increase of progressive assimilation with increasing pitch (Table II). This increase cannot be due to speech tempo; that would contradict previously obtained results of measurements on the influence of speaking rate, which showed an increase in regressive assimilation in stop-stop clusters (Slis, 1981a). Consequently, speaking rate can be excluded as the origin for the increase in progressive assimilation, and we have to attribute this to monotonisation and to a rise in pitch. These results corroborate our hypothesis that an increase of stiffness of the vocal cords with an increase of pitch cause less favourable boundary conditions for phonation, thus leading to more progressive assimilation. The observed difference between normally intonated speech and speech on a monotone can be explained by assuming that the subjectively experienced extra effort needed to maintain a monotone is attended by a greater tension (larger stiffness) of the internal larynx musculature.

Comparison between good and poor voices without training shows significantly more progressive (less regressive) assimilation in poor than in good voices; trained poor voices occupy an intermediate position (Table IV).

The acoustic measurements show a significant difference between the spectra of good and poor voices (see above). It seems probable that the organic differences that lead to these spectral differences also cause the

Table IV. Frequency of assimilation of voice as a function of voice quality. In parentheses: frequency based on chance.

	good voices (N=11)	poor voices trained (N=11)	poor voices untrained (N=12)
No assimilation	92 (93.7)	77 (93.9)	121 (102.4)
Regressive ass.	364 (322.5)	336 (323.1)	298 (352.5)
Progressive ass.	71 (110.8)	115 (111.0)	157 (121.1)

$\chi^2 = 45.79$, df=4, $p < .001$.

differences in assimilation. For the same reason as given for differences observed with different pitches, differences in direction of assimilation between good and poor voices cannot be attributed to differences in speech tempo.

No spectral difference can be observed between trained and untrained poor voices. This makes good sense, since it seems highly improbable that the organic condition can be changed by only a few training sessions. Yet, the frequency of assimilation in trained poor voices is intermediate between good and untrained poor voices. We conclude therefore that the increase of regressive and the decrease of progressive assimilation in trained vs. untrained voices is not due to an improved organic condition, but must be caused by another mode of voicing. In both the trained and the untrained group, nine of the subjects were judged to speak habitually on too high a pitch. The speech therapist stated that after a few instructions most subjects were able to keep their pitch at a good, normal level. Since comparison between speech at a high pitch with that at a low pitch shows more regressive assimilation at the lower pitch (see above), we conclude that assimilation changed because of a change in pitch by training.

In the speech of trained speakers with poor voices we observe a relatively high frequency of assimilation. This may be caused by a higher speech tempo than in the speech of the other groups of speakers (cluster duration is shorter: Tables V & VI) or it may be the result of speech training; in some of the exercises the attention of the trainees is directed towards a continuous speech flow (viz. in "resonance" exercises).

Table V. Mean cluster duration (in ms) as a function of voice quality and assimilation. In parentheses: standard deviation (in ms).

	good voices	poor trained voices	poor untrained voices
No assimilation	238 (63)	207 (54)	211 (57)
Regressive ass.	164 (32)	154 (33)	151 (32)
Progressive ass.	176 (46)	148 (30)	172 (47)

Table VI. Differences in the cluster duration of table V (in ms).

	good vs. untrained poor voices	good vs. trained poor voices	untrained poor vs. trained poor voices
No assimilation	27 *	31 **	4
Regressive ass.	13 **	10 **	-3
Progressive ass.	3	27 **	24 **

* p. < .01 (t-test)
** p. < .001 (t-test)

Since no diagnosis was made with respect to the origin of the poor voice quality of our speakers, we do not know whether the two groups with poor voices form homogeneous groups. In fact we expect that the causes of their poor voice quality are multiple. Therefore we can only give a general description of the mechanism we assume to give rise to the phenomena observed. We hypothesize that in poor voices the boundary conditions for vocal fold vibration are less favourable than in good voices. It is possible that the airstream cannot be used optimally because of an irregular surface of the vocal folds or because of too high a tension of the larynx musculature leading to increased stiffness. Irrespective of the cause, voicing is expected to be interrupted earlier in poor than in good voices with a low rate of airflow, as is the case during consonant obstructions. In cases where the interruption of voicing crosses the boundary between regressive and progressive assimilation, viz. by shortening the voice tail and/or delaying VOT, cases of regressive assimilation change into progressive assimilation.

7. CONCLUSION

Our hypothesis that speech of speakers with poor voices and speech at a high pitch will show more progressive assimilation than speech of speakers with good voices and speech at a low pitch is confirmed by the results. We therefore conclude that organic differences and differences in the mode of phonation have an influence on assimilation of voice. Moreover, assuming that the vibration condition of the vocal cords in good voices (or at a low pitch) is better than in poor voices (or at a high pitch), we conclude that a better vibration condition leads to more regressive assimilation.

A sociolinguistic explanation as was forwarded with respect to differences between male and female voices is highly improbable. Since the organic condition as such is found to influence assimilation of voice in good and poor voices, we think it justifiable to assume that differences in assimilation due to the sex of the speakers are also primarilv caused by organic differences.

ACKNOWLEDGEMENTS

1. Above all I am indebted to Toon Cohen for the stimulating way he introduced me to the field of phonetics. With respect to the present experiments, I want to thank Renée van Bezooijen, Loe Boves, Bert Cranen, Frans Decates and Toni Rietveld for the support they offered in the various stages of the experiments. Paul Bovy and Jan Buiting assisted in collecting and processing material as part of their study in linguistics.

REFERENCES

Boves, L. & Cranen, B. (1982) On long-term average spectra and their representation *Proceedings of the Institute of Phonetics Nijmegen*, **6**, 20-30.

Erp, A. van, Halfens, Y.M A., Hoogeveen, W H.J.J. & Zink, A H M (1981). The effect of stress on assimilation of voice in Dutch and English *Proceedings of the Institute of Phonetics Nijmegen*, **5**, 75-83.

Halle, M. & Stevens, K.N. (1971). A note on laryngeal features. *M.I.T. Quarterly Progress Report*, **101**, 198-213

IBM SSP (1970). *System/360 scientific subroutine package, version III, program number 360A-CM-03X* IBM Corp., Techn Publ. Dept.

Luchsinger, R. & Arnold, G E. (1965) *Voice-Speech-Language, Clinical Communicology Its Physiology and Pathology* Wadsworth, Belmont, Cal.

Overall, J E. & Klett, C.J. (1972). *Applied multivariate analysis*. McGraw Hill, New York

Rossum, N van (1981) Parallel power density measurement system *Proceedings of the Institute of Phonetics Nijmegen*, **5**, 142-159.

Slis, I H. (1970) Articulatory measurements on voiced, voiceless and nasal consonants: A test of a model. *Phonetica*, **21**, 193-210.

Slis, I H (1972) On coarticulation of articulatory effort and syllable boundaries. *IPO Annual Progress Report*, **7**, 40-44

Slis, I.H. (1981a). The effect of speaking rate on assimilation of voice. *Proceedings of the Institute of Phonetics Nijmegen*, 5, 160-176.

Slis, I H (1981b) The influence of the position of a stressed syllable on assimilation of voice *Proceedings of the Institute of Phonetics Nijmegen*, 5, 177-187.

Slis, I H. (1981c) Half a dozen influences on assimilation of voice in Dutch. *Proceedings of the Institute of Phonetics Nijmegen*, 5, 188-203.

Slis, I H (1981d). Reflections on a theoretical frame for the assimilation of voice. *Proceedings of the Institute of Phonetics Nijmegen*, 5, 204-215.

CHAPTER 8

RULES FOR ASSIMILATION OF VOICE IN DUTCH

I. H. Slis

Accepted for publication in:

IN HONOUR OF ILSE LEHISTE / ILSE LEHISTE PÜHENDUSTEOTS
Edited by R. Channon and L. Shockey
Dordrecht: Foris Publ., 1985

RULES FOR ASSIMILATION OF VOICE

INTRODUCTION

A tremendous amount of research has been done on single voiced and voiceless consonants. Combinations of these consonants in clusters are less thoroughly studied. Previously, however, in the "pre-instrumental times" of phonetic research, assimilation of voice was a topic among Dutch phoneticians. The views on the subject varied very much and were sometimes even conflicting. For instance, Van Ginneken (1935) stated that assimilation of voice was largely regressive because of foreign influences; this regressive assimilation had so much in common with pre-slavic regressive assimilation that this could not be accidental. On the other hand, Pée (1948) concluded on the basis of a large amount of material that progressive assimilation is nearly always the rule in Dutch, although not carried through equally in all combinations. Demeulemeester (1962) was of the opinion that assimilation was not regular; he observed that "...with radio broadcasters a specific system cannot be found for the direction their assimilations take..." (1962:31, my translation).

RULES FOR THE ASSIMILATION OF VOICE

Despite unpredictable assimilations, as observed by Demeulemeester, and differences of opinion, for instance Van Ginneken versus Pée, most investigators agreed on some general tendencies. Two main rules can be formulated for clusters consisting of plosives and fricatives:

(1) Regressive assimilation is observed in the majority of two-consonant clusters if the second consonant is a voiced plosive;

(2) Progressive assimilation is generally found if the second consonant is a voiced fricative.

A third rule that is relevant states that:

(3) All syllable final plosives and fricatives are devoiced in Dutch.

Therefore, the cases of assimilation of voice under investigation consist of a voiceless plosive or fricative followed by a voiced plosive or fricative.

Practical all investigators in the field observed a number of deviations from the rules; many of these deviations could be attributed to the dialectal backgrounds of the investigators and their subjects. Moreover, a great number of other possible influences were assumed. A survey of all the factors that were suggested are listed below in random order:

1. Phonological composition of the utterance (e.g. Cohen et al. 1971)
2. Frequency of occurence of the speech sounds used (Meinsma 1958)
3. Spoonerisms (Meinsma 1958)
4. Etymology (e.g. Van Ginneken 1935, Leenen 1954)
5. Orthography (e.g. Eijkman 1933)
6. Word accent (e.g. Eijkman 1933, Van Haeringen 1955)
7. Semantic value of words (Van Haeringen 1955)
8. Word boundaries (e.g. Meinsma 1958)

9. Type of text (e.g. colloquial, official,...)(Leenen 1954)
10. Emotion (Van Ginneken 1935, Kaiser 1942, Meinsma 1958)
11. Speech rate (e.g. Kaiser 1942 and 1958, Meinsma 1958)
12. Experimental situation (e.g. Meinsma 1958, Pée 1948)
13. Sex of the speaker (Demeulemeester 1962, Kaiser 1958)
14. Dialect (e.g. Demeulemeester 1962, Eijkman 1933, Van Ginneken 1935, Kaiser 1958, Meinsma 1958)
15. Perceptual errors of the investigators (Meinsma 1958)

RULES BASED ON TWO FEATURES

The large number of suggested influences illustrates the complexity of the problem. Therefore it is not surprising that in generative phonology complex rules are used to describe assimilation of voice. In addition to the distinctive feature "voice", a second feature is introduced. In this context we find the features "foreign" (Mey 1968), "pause" (Tops 1974 referring to Brink 1970), and "tense" (e.g. Hubers and Kooij 1973). Among phoneticians we also find supporters of two independent features to describe the voiced-voiceless distinction (e.g. Debrock 1977, Kim 1965). In addition to the voicing feature they assume a feature incorporating articulatory effort, e.g. "tense" or "fortis". Others prefer one underlying feature which can be reflected in a number of articulatory and acoustic parameters (e.g. Slis and Cohen 1969, Slis 1970).

Among Dutch phonologists a two-feature description seems to be most common. In the grammar proposed by Hubers and Kooij (1973) a final devoicing rule is hypothesized, which operates in three steps:

(a) $|\text{Obs}| \rightarrow |+\ \text{Tns}|\ /\ \underline{\qquad}\ \#$

(b) $|\text{Obs}| \rightarrow |-\ \text{Vce}|\ /\ |\overline{+\ \text{Tns}}|$

(c) $|\text{Obs}| \rightarrow |-\alpha\ \text{Vce}|\ /\ |\overline{\alpha\ \text{Tns}}|$

These rules change all final obstruents into voiceless, tense obstruents. This corresponds to the general tendency we mentioned under (3) above.
A second set of three operations concerns assimilation:

(d) $$|\text{Obs}| \rightarrow |+\ \text{Vce}|\ /\ \underline{\qquad}\ \#\ \left|\begin{matrix}\text{Obs} \\ -\ \text{Cnt} \\ +\ \text{Vce}\end{matrix}\right|$$

This operation voices all obstruents before voiced plosives and is, in fact, regressive assimilation of voice before voiced plosives (general tendency (1), above).

(e) $$|\text{Obs}| \rightarrow |-\ \text{Vce}|\ /\ \left|\begin{matrix}\text{Obs} \\ \\ -\ \text{Vce}\end{matrix}\right|\ \#\ \underline{\qquad}$$

This operation devoices all obstruents after voiceless obstruents. Since in operation (d) all obstruents before plosives assume [+ voice], only obstruents before fricatives remain voiceless. Consequently, step (e) describes progressive assimilation before fricatives (general tendency (2)).

$$\text{(f)} \quad |\text{Obs}| \rightarrow |+ \ \text{Tns}| \ / \ \begin{vmatrix} \text{Obs} \\ \\ + \ \text{Tns} \end{vmatrix} \ \# \ \_\_\_\_$$

In this operation all obstruents after [+ Tense] obstruents become [+ Tense]. Since in the two preceding assimilation rules the [+ Tense] character of the first consonant is not changed, all second consonants become [+ Tense]. This results in a cluster of two [+ Tense, - Voice] consonants if the second consonant is a fricative. These can be regarded as two normal voiceless consonants and therefore progressive assimilation takes place. If the second consonant is a plosive, however, the rules yield two [+ Tense, + Voice] consonants. Therefore regressive assimilation takes place with respect to voicing and progressive assimilation with respect to tenseness. The feature [+ Voice] promotes a voiced realization of the cluster and the feature [+ Tense] a voiceless realization. In this way an explanation is provided for the observed inconsistencies in assimilation before stops.

A shortcoming of these rules is that in fact they generate question marks in the latter case since they do not predict under what conditions the ambivalent [+ Tense, + Voice] result tends towards a voiced and when towards a voiceless realization. The literature indicated that deviations are to a certain extent predictable. For instance Leenen (1954) observed that a group of function words beginning with a /d/ led to progressive assimilation instead of regressive assimilation. It has been suggested that for exeptions like this special rules have to be added. One may wonder whether for each of the possible influences that are suggested in the literature, special rules ought to be introduced. This would lead to a very complex system. Moreover, a grammar based on binary features and unequivocal rules will never lead to inconsistencies as observed by Demeulemeester (1962) or to different grades of partial assimilation as observed by e.g. Van Rijnbach and Kramer (1939).

We think that a better description can be obtained by introducing gradual features such as proposed by Ladefoged (1975) and variable rules of the kind proposed by Labov (1972) for sociolinguistic problems. We will illustrate this by means of the results of an experiment in which we investigate the influence of speech rate on assimilation of voice.

EXPERIMENT: ASSIMILATION AS FUNCTION OF SPEECH RATE

The original aim of our experiment was to investigate whether the subjective data we found in the literature could be confirmed by objective measurement data. For this purpose we needed objective criteria which will be described below. In addition, we wanted to examine to what extent some of the factors mentioned play a role in changing the direction of assimilation. In the present experiment we first looked at the speech rate. In addition, we examined influences of the phonological composition of the clusters and of the sex and region of origin of our speakers.

Ten different sentences were used, containing thirteen obstruent-obstruent clusters in a pre-stress position. Four of the clusters contained two stops (/pb,tb,pd,td/), four a fricative followed by a stop (2x /fb/, /sb,fd/), three a stop followed by a fricative (2x /tv/, /pz/) and two contained two fricatives (/sv,fz/).

Twenty-five subjects were instructed to read these ten sentences aloud first slowly, then at a normal speed and finally as fast as they could. The 25 subjects (students in the linguistics departments of Utrecht and Nijmegen) were not informed about the aims of the experiment. All subjects spoke standard Dutch although traces of their local pronunciation were observable. Among the subjects, eight men and eight women were from the same regional background; their speech was used to study the influence of sex on assimilation. Besides we could compare a group of four men and four women from Limburg (south-east) with a similar group from the middle-east region of the country, and a group of seven men from Holland (west) with eight men from "other regions" (middle and south-east).

OBJECTIVE CRITERIA

We based our criteria on measurements on single voiced and voiceless stops. In intervocalic voiceless stops voice activity continued till about 10-20 ms (s.d. 10 ms) after the moment of oral closure (Slis 1968 and 1970). We assumed that a continuation of voicing, which we called "voice tail", between zero and 50 ms is normal in voiceless obstruents. If the voice tail exceeded 50 ms, we took it that the post-vocal obstruent became voiced by assimilation.

In initial prevocalic voiced stops we found a voice onset before the moment of oral release, in other words a negative voice onset time (V.O.T.). With initial and intervocalic voiceless stops the voice onset occurred after oral release. Klatt (1975) observed for English that in clusters with a voiceless fricative followed by a voiceless stop, V.O.T. is advanced to the moment of opening. Although we did not examine this for Dutch, we accidentally observed that the same holds for Dutch. We therefore defined the second obstruent voiced if the V.O.T. is negative and voiceless if the V.O.T. is zero or positive.

One additional remark has to be made. We observed clusters in which voicing activity started during the closed interval after a preceding voiceless obstruent and in which voicing was interrupted a second time before oral disclosure. We hypothesized that this second voiceless interval was due to aerodynamic conditions and that this second interruption was meant to be voiced. We defined the first time the voicing started during the closed interval to be the V.O.T. and ignored the second interruption. A similar reasoning held for clusters in which voicing was interrupted after a voice tail longer than 50 ms; in these cases we assumed that voicing was intended to continue and we regarded these as cases of regressive assimilation.

Summarizing, we arrived at the following definitions for three categories of assimilation (see also fig. 1):

1. No assimilation: The first consonant is voiceless and the second voiced. The voice tail is shorter than 50 ms and the V.O.T. is negative. Voicing may be interrupted after the V.O.T.; this interruption is ignored.
2. Regressive assimilation: Both consonants are voiced. In most cases voicing continues during the closed interval. Voicing may be interrupted after a voice tail longer than 50 ms; this interruption is regarded as being due to aerodynamic conditions and will be ignored.
3. Progressive assimilation: Both consonants are voiceless. The voice tail is shorter than 50 ms and V.O.T. is zero or positive.

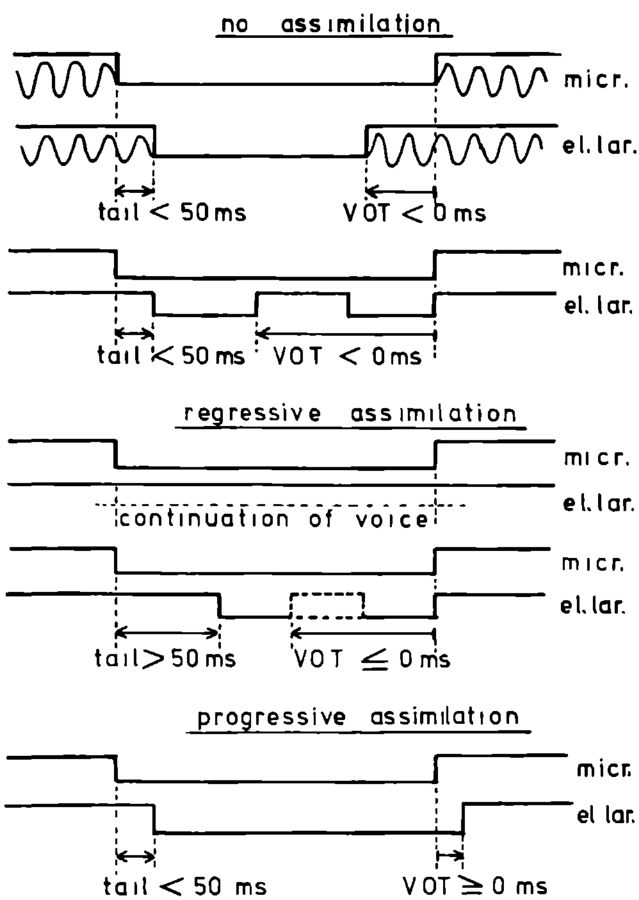


fig. 1 Schematic representation of the definitions of "no assimilation", "regressive assimilation" and "progressive assimilation"

Although slightly different conditions for our definitions can be thought of, the ones chosen proved to be satisfactory since they are easy to use; only in a few cases was there uncertainty about the right categorization. Also from the perceptual point of view objections can be raised. It is possible that a number of listeners will perceive a voiced consonant if V.O.T. is zero. However, a V.O.T. of zero is relatively easy to measure and, moreover, it is impossible to decide on a positive value for V.O.T. that meets all requirements.

MEASUREMENTS

For the sake of the determination of voice tails and V.O.T.s it was necessary to measure the moments of oral closing and opening and of the end and beginning of voice activity.
The moments of oral closing and opening were derived from UV-oscillograms of the acoustic speech signal. The moments of closing were defined at the sudden decrease of the amplitude at the end of the preceding vowel. With fricatives and voiced plosives the moments of opening were defined at the points of sudden increase of the amplitude and with voiceless plosives at the point of the noise burst.
Voicing was determined by simultaneously recording the output of an electroglottograph (Fourcin and Abberton 1971). The end and beginning of voice activity were defined as the end and beginning of the visible oscillations of the trace. In order to find the voice tails and V.O.T.s the "voice signal" was compared directly with the "speech signal". Using these comparisons, we observed that in a few cases voicing continued in the speech signal trace while it was interrupted in the voice signal. We hypothesized that in these instances the vocal cords vibrated without touching each other. We decided that these intervals had to be regarded as voiceless since the speaker tried to produce a voiceless consonant by opening his glottis. Presumably, vibration of the edges of the vocal cords was sustained by aerodynamic conditions in these cases.

In the processing of the data we made use of the chi-square test. If in one of the categories the frequency was too low (less than 5) we used the L-test (Spitz 1968). The results are given in Tables I-VI. After the frequencies of occurrence we indicated, between brackets, the frequency which would reflect a chance level.

RESULTS: PHONOLOGICAL COMPOSITION OF THE CLUSTER

Highly significant differences are found between clusters ending with a stop and those ending with a fricative (Table I). If the second consonant is a fricative, 100 % of the assimilations are progressive. If this is a stop we observe that regressive assimilation is about three times as frequent as progressive assimilation.

Table I Frequency of assimilation as a function of the manner of articulation of the second consonant (= C2). Between brackets: Frequency based on chance

	C2 = stop	C2 = fricative	
No assimilation	170 (125)	34 (79)	L = 322.14
Regressive ass.	329 (202)	0 (127)	df = 2
progressive ass.	99 (270)	341 (170)	p < .001

In Table II we divided up the obstruent-stop cluster data based on the first consonant. Again we find a highly siginificant difference. In stop-stop clusters 12 % of the assimilations are progressive and in fricative-stop clusters 37 %.

Table II Frequency of assimilation as a function of the first consonant (= C1) in obstruent-stop clusters. Between brackets: Frequency based on chance

	C1 = stop	C1 = fricative	
No assimilation	63 (85)	107 (85)	chi sq. = 53.07
Regressive ass.	208 (165)	121 (165)	df = 2
Progressive ass.	28 (50)	71 (50)	p < .001

These two comparisons show that the presence of a fricative seems to lead to more progressive assimilation. We feel that this is in line with the observation that fricatives have a tendency to become voiceless in Dutch (Cohen et al. 1971).

Table III Frequency of assimilation as a function of the place of articulation of the second consonant (= C2) in stop-stop clusters. Between brackets: Frequency based on chance

	C2 = /b/	C2 = /d/	
No assimilation	32 (31)	31 (32)	chi sq. = 5.64
Regressive ass.	109 (104)	99 (104)	df = 2
Progressive ass.	8 (14)	20 (14)	.05 < p < .10

We also investigated whether the place of articulation of a stop in second position in the cluster plays a role. A tendency towards more progressive assimilation with alveolar /d/ than with labial /b/ can be observed (Table III). The number of progressive assimilations in our material is too low to obtain any significant results. If we add similar data from other experiments (Partly published in Slis 1981 b) the difference becomes highly significant (p < .001). Consequently, we may assume that the influence of word initial /d/ (Leenen 1954) is not restricted to a group of function words.

RESULTS: SPEECH RATE

Since the phonological composition has consequences for assimilation of voice we divided our results into four groups, viz. stop-stop, fricative-stop, stop-fricative and fricative-fricative clusters (Table IV).

Table IV Frequency of assimilation as a function of speech rate. Between brackets: Frequency based on chance

stop-stop	slow	normal	fast	
No assimilation	43 (21)	16 (21)	4 (21)	L = 64.96
Regressive ass.	44 (70)	72 (70)	92 (69)	df = 4
Progressive ass.	13 (9)	12 (9)	3 (9)	$p < .001$
Fricative-stop				
No assimilation	54 (35)	36 (36)	17 (36)	chi sq. = 30.78
Regressive ass.	30 (40)	39 (40)	52 (40)	df = 4
Progressive ass.	15 (23)	25 (24)	31 (24)	$p < .001$
stop-fricative				
No assimilation	10 (4)	1 (4)	0 (4)	L = 18.32
Regressive ass.	0 (0)	0 (0)	0 (0)	df = 2
Progressive ass.	65 (71)	74 (71)	75 (71)	$p < .001$
fricative-fricative				
No assimilation	17 (8)	6 (8)	0 (8)	L = 27.72
Regressive ass.	0 (0)	0 (0)	0 (0)	df = 2
Progressive ass.	33 (42)	44 (42)	50 (42)	$p < .001$

Speech rate has a significant influence on assimilation; with increasing rate the number of assimilated clusters increases in all four groups. If the second consonant is a fricative the increase is progressive. In stop-stop clusters regressive assimilation increases from 77 % in slow towards 96 % in fast speech. The few remaining cases of progressive assimilation in fast speech concern clusters in which the second consonant is an alveolar /d/.
In fricative-stop clusters the increase is distributed proportionally between regressive and progressive assimilation.

These results suggest that with a higher speech rate assimilation increases in the usual direction of the cluster, viz. progressive for stop-fricative and fricative-fricative clusters and regressive in stop-stop clusters. In fricative-stop clusters the first consonant, a fricative, promotes progressive assimilation and the second consonant, a stop, regressive assimilation. These opposite tendencies seem to be of about equal magnitude resulting in an equal increase in both directions.

RESULTS: SEX OF THE SPEAKER

We restricted ourselves to clusters in which the second consonant is a stop, since in those clusters assimilation occurs in both directions. The results show a highly significant difference between male and female speakers (Table V). Women show a higher degree of assimilation (76 %) than men (68 %), and more progressive assimilation is found in women (35 %) than in men (12 %). Both differences are in agreement with the literature (Demeulemeester 1962, Kaiser 1958).

Table V Frequency of assimilation as a function of sex of the speaker in obstruent-stop clusters. Between brackets: Frequency based on chance

	8 male speakers	8 female speakers	
No assimilation	62 (54)	46 (54)	chi sq. = 22.77
Regressive ass.	114 (104)	93 (103)	df = 2
Progressive ass.	16 (34)	51 (33)	$p < .001$

We propose two different kinds of explanations, viz. a mechanical and a sociolinguistic one:

1. A mechanical explanation: There are substantial anatomical differences between the vocal organs of men and women. The vocal cords of men are longer and heavier than those of women. Therefore, the conditions for vocal fold vibration are different. For instance the higher pitch of female voices can be traced back to anatomical differences. Against this background it should be highly surprising if the boundary conditions for vocal cord vibration were not different. We suggest that the vocal fold vibrations are easier to sustain for men than for women, thus leading to more voicing and consequently to more regressive assimilation in men.
2. A sociolinguistic explanation: A number of cases are known in which the pronunciation of men and women differ. It is generally accepted that women use "better", "more prestigious" or "more correct" speech (e.g. Smith 1979, Trudgill 1974). With respect to assimilation of voice, Gussenhoven (1981) found that women are inclined to judge progressive assimilation more favourably than men. The fact that the majority of linguists indicate that regressive assimilation is correct in obstuent-stop clusters and that women display a higher proportion of progressive assimilation than men might indicate that a development towards progressive assimilation of voice is taking place in Dutch.

RESULTS: REGION OF ORIGIN

In order to compare the speech of subjects originating from different dialect regions only obstruent-stop clusters were used, again with view to the fact that assimilation could occur in both directions.

Kohler (1979) related progressive assimilation in Germanic languages to the aspiration of voiceless plosives, and regressive assimilation in Romanic languages to the absence of aspiration. Since the eastern dialects of The Netherlands have aspiration as opposed to the southern dialects we expected similar differences in assimilation between people from Twente and de Achterhoek (east) and from Limburg (south-east). A group of eight speakers from the east, four male and four female, was compared with a similar group from the south-east. No siginificant differences were observed (chi square = 2.54, df = 2, $.2 < p < .3$).

A second comparison was made between the speech of seven male speakers from Holland (west) and eight male speakers from the east and south-east. It is generally assumed that the western people show more progressive assimilation. Again, we did not find significant differences (chi square = 1.29, df = 2, $.5 < p < .7$).

The explanatory value of a different dialectal background to interpret differences in results as was done often in the older literature is rather questionable in the light of our results. However, we used the standard Dutch pronunciation of students in linguistics. It may well be that the differences mentioned in the literature can be found in true dialect speech.

INCONSISTENCIES WITHIN SPEAKERS

If we think in terms of a set of strict rules that predict a given direction of assimilation under fixed conditions (which may take context, situation or speaker into consideration), we would expect the same direction of assimilation in repetitions of a given utterance by one speaker. In our experiment, we recorded every sentence spoken three times by each subject, the only difference in the condition being the speech rate.
The results of the experiments showed that an increase in rate resulted in a greater number of assimilated clusters. The direction of assimilation increased in the preferred direction for the cluster type. Only in stop-stop clusters we observed a change in direction; thirteen progressive assimilations in slow speech changed into three in fast speech. Consequently, ten clusters with progressive assimilation changed into regressive assimilation.
A change of "no assimilation" with slow speech into progressive or regressive was not regarded as "inconsistent"; this change was due to speech rate. We restricted our definition of inconsistent assimilation to cases in which one speaker showed progressive assimilation in the production of one cluster in one tempo and regressive assimilation in

another tempo.
Since 25 subjects participated and the sentences contained eight obstruent-stop clusters, the subjects had the possibility of being inconsistent 200 times. In 48 of these cases, two or three of the realisations showed no assimilation. Therefore, 152 possibilities remained. We counted 35 (= 23 %) inconsistencies in 16 out of 25 speakers. This was 25 more then the ten inconsistencies that were due to increase in tempo of stop-stop clusters. On this basis we concluded that our subjects did not use strict rules. Speakers seemed to aim at a norm distribution of the direction of assimilation. Such behaviour can be described by the variable rules which Labov (1972) uses for sociolinguistic descriptions.

PARTIAL ASSIMILATION

According to our definition of regressive assimilation, it is not necessary that voicing be present throughout the entire cluster. In fact we indicated cases in which the voice activity can be interrupted. We classify these as clusters with partial regressive assimilation. Clusters in which voicing continues without interruption are here classified as total regressive assimilation.
In practice this criterion proved to be difficult to use in a few cases. For instance, with poor voices it was difficult to decide whether the vocal cords vibrated irregularly, or whether the voice was interrupted during one period. In some other cases the amplitude of the larynx signal gradually approached zero, indicating incomplete closure of the glottis during vibration. We tried to maintain a fixed criterion for these complicated cases.

A large part of the clusters that we categorized as regressive assimilation proved to be partially assimilated, viz. 45 % in stop-stop and 30 % in fricative-stop clusters. In female speakers regressive assimilation was more frequently partial (63 %) than in male speakers (25 %). This increased the differences in assimilation in male and female speech (Table VI).

Table VI Frequency of total and partial regressive assimilation and progressive assimilation as a function of sex of the speakers in obstruent-stop clusters. Between brackets: Frequency based on chance

	8 men	8 women	
Total regressive ass.	86 (57)	35 (64)	chi sq. = 49.66
Partial regress. ass.	28 (41)	58 (45)	df = 2
Progressive ass.	16 (32)	51 (35)	p < .001

From these measurements we conclude that a neat, discrete classification for assimilation is impossible. The criteria will always be arbitrary. Depending on the choice of the criteria, the outcomes will be different.

This is shown in the comparison of total and partial regressive assimilation between men and women. The more or less continuous scale of the degree of assimilation we find in our material was also observed by Van Rijnbach and Kramer (1939). In our opinion this gradual phenomenon can best be described by means of a grammar with gradual features (Ladefoged 1975). The value of the features can be dependent on the components of the cluster.

Our present results show that fricatives promote progressive assimilation. The feature "voice" in fricatives must therefore have a more voiceless value on the scale than in stops. Our former observation (Slis 1970) that the glottis opens wider in voiceless fricatives than in voiceless plosives, and also the observation by Cohen et al. (1971) that initial voiced fricatives tend to be realized as voiceless in Dutch, support this view.

DISCUSSION AND CONCLUSION

From our results we conclude that the two major rules for assimilation of voice, viz. regressive assimilation before stops and progressive assimilation before fricatives, describe only the main tendencies. In the present experiment assimilation is always progressive before fricatives. In another experiment (Slis 1981 a), however, we observed a few cases of regressive assimilation before fricatives. With regard to assimilation before stops, we only find 75 % regressive assimilation. Since in obstruent-stop clusters regressive as well as progressive assimilation occurs, we can use these clusters to investigate the effect of experimental conditions on the direction of assimilation.
We find that the presence of a fricative in the cluster shifts the ratio between progressive and regressive assimilation towards more progressive assimilation. The same is observed with alveolar /d/ vs. labial /b/. A shift of this type suggests competitive forces that are not exactly equal; fricatives and alveolar stops pull the results towards voiceless, and must have more voiceless power. The shift due to fricatives is not equal to the shift due to alveolar stops. It is impossible to describe these effects with binary features.
Moreover, we observe a large number of partial assimilations. These show large variations with respect to the amount of "voicelessness". As we argued above, these can be described by gradual features. The value on the feature scale indicates the magnitude of the competitive force.

In addition to these phonological influences, the results show differences in assimilation between and within subjects. For example we observe more progressive assimilation with female than with male speakers. Moreover, the subjects are inconsistent; in analogous situations they use progressive as well as regressive assimilation. The inconsistencies can not be described by strict rules. Since the cause of the differences presumably lies outside the strict field of linguistics (we suggested either a mechanical cause or a preference of the speaker), we think it justifiable to describe these observations by variable rules.

In a considerable number of clusters we find no assimilation. This can be accounted for by optional rules. The degree to which the rules are optional depends on different factors, for instance speech rate or sex of the speaker. These factors presumably imply mechanical constraints (viz. time available for alternating successive articulatory gestures) or sociolinguistic drives (preference of the speaker). We therefore propose, as with subject dependent influences, variable controls for optional rules.

In a former paper we put forward the hypothesis that the voiced-voiceless distinction in Dutch is synonymous with the tense-lax distinction (Slis 1971). The labels "voice" and "effort" indicate different aspects of the same distinction. In our model more measurable aspects joined in one complex. In our present experiment we come to the conclusion that we do need gradual features and variable rules to describe assimilation of voice. On the other hand, we do not find the need for two independent features, as is assumed in phonology. Therefore, we maintain our previous conclusion that the voiced-voiceless distinction can be described by one feature in Dutch. This also holds for the description of assimilation.

Acknowledgements: I want to thank Wim Hildering, Nanda Poulisse and Els Spoeltman who assisted in collecting and processing the material as a part of their study in linguistics. The staff of the Institute of Phonetics of the University of Utrecht helped me in finding subjects from the western regions of the country. Alda van Erp was a great help in correcting the English text.

REFERENCES

Brink, D., Problems in phonological theory: A generative phonology of Dutch (Ph.D.-thesis, University of Wisconsin). Ann Arbor: University Microfilms, 1970

Cohen, A., Ebeling, C.L., Fokkema, K., Holk, A.G.F. van, Fonologie van het Nederlands en het Fries, 2-nd edition, Chapter II-16: Assimilatie en sandhi, Chapter IV-5: Classificatie van medeklinkers, and Chapter V-18: Assimilatie, sandhi. 's Gravenhage: Martinus Nijhoff, 1971, 50-53, 101-103, and 129-131

Debrock, M., An acoustic correlate of the force of articulation. Journal of Phonetics 5, 1977, 61-80

Demeulemeester, F., Assimilatie van stemloze eksplosieven en frikatieven voor B en D. Taal en Tongval 14, 1962, 20-36

Eijkman, L.P.H., Assimilatie. Nieuwe Taalgids 27, 1933, 315-325

Fourcin, A.J., Abberton, E., First applications of a new laryngograph. Medical and Biological Illustration 21, 1971, 172-182

Ginneken, J. van, Ras en Taal, 33: de Nederlandse sandhi, 57:

silbenvorming en sandhi. Amsterdam: Verhandelingen van de Koninklijke Akademie van Wetenschappen, afd. Letterkunde, Nieuwe Reeks 36, 1935, 100-108 and 143-148

Gussenhoven, C., Measuring the acceptability of voiced fricatives in Dutch. Proceedings Institute of Phonetics Nijmegen 5, 1981, 96-129

Haeringen, C.B. van, Is tat juist, op tie manier. Taal en Tongval 7, 1955, 49-57

Hubers, G.A.C., Kooij, J.G., Voice assimilation in Dutch. Acta Linguistica Hafniensa 14, 1973, 25-33

Kaiser, L., Biological and statistical research concerning the speech of 216 Dutch students, IV. Archives Néerlandaises de Phonétique Expérimentale 18, 1942, 1-58

Kaiser, L., Onderzoek naar assimilatieverschijnselen, in: Album Edgar Blanquaert. Tongeren: George Michels, 1958, 31-41

Kim, C.W., On the autonomy of the tensity feature in stop classification. Word 21, 1965, 339-359

Klatt, D.H., Voice onset time, frication, and aspiration in word initial consonant clusters. Journal of Speech and Hearing Research 18, 1975, 686-706

Kohler, K.J., Parameters in the production and the perception of plosives in German and French. Arbeitsberichte Institut für Phonetik Kiel 12, 1979, 261-291

Labov, W., Sociolinguistic Patterns. University of Pennsylvania Press, 1972

Ladefoged, P., A course in Phonetics. New York: Harcourt, Brace, and Jovanovich, 1975

Leenen, P., Een Limburgse en Nederlandse uitspraakregel. Taal en Tongval 6, 1954, 1-24

Meinsma, G.L., Assimilatie en assimilatieonderzoek. Taal en Tongval 10, 1958, 107-114

Mey, J., A case of assimilation in modern Dutch. Acta Linguistica Hafniensa 11, 1968, 123-145

Pée, W., Regressieve of progressieve assimilatie, in: Miscellania J. Gessler, bibliographia Gessleriana (Roemans, R., Ed.). Antwerpen, 1948, 979-988

Rijnbach, A. van, Kramer, W.F., De assimilatie van stem en plaats in het Nederlands bij het lezen van losse zinnen. Onze Taaltuin 8, 1939, 97-107

Slis, I.H., Experiments on consonant duration related to the time structure of isolated words. IPO Annual Progress Report 3, 1968, 71-80

Slis. I.H., Articulatory measurements on voiced, voiceless and nasal consonants: A test of a model. Phonetica 21, 1970, 193-210

Slis, I.H., Articulatory effort and its durational and electromyographic correlates. Phonetica 23, 1971, 171-188

Slis, I.H., The influence of the position of a stressed syllable on assimilation of voice. Proceedings Institute of Phonetics Nijmegen 5, 1981 a, 177-187

Slis, I.H., Half a dozen influences on assimilation of voice in Dutch. Proceedings of the Institute of Phonetics Nijmegen 5, 1981 b, 188-203

Slis, I.H., Cohen, A., On the complex regulating the voiced-voiceless distinction I and II. Language and Speech 12, 1969, 80-102 and 137-155

Smith, P.M., Sex markers in speech, in: Social markers in speech (Scherer, K.R., Giles, E., Eds). Cambridge: Cambridge University Press, 1979, 109-146

Spitz, J.C., Statistiek voor psychologen, pedagogen, sociologen, 3-rd edition. Amsterdam: Noord-Hollandse Uitg. Mij., 1968, 354-357

Tops, G.A.J., Assimilation of voice: A generative approach. Leuvense Bijdragen 63, 1974, 43-51

Trudgill, P., Sociolinguistics: An Introduction. Harmondsworth: Penguin Books Ltd, 1974

CHAPTER 9

ASSIMILATION OF VOICE AS A FUNCTION OF STRESS AND WORD BOUNDARIES

I. H. Slis

Reprint from:

PROCEEDINGS OF THE INSTITUTE OF PHONETICS NIJMEGEN
No. 7, 1983, 74-82

ASSIMILATION OF VOICE AS A FUNCTION OF STRESS AND WORD BOUNDARY

1. INTRODUCTION

In a previous paper (Slis 1981 d) we provided a theoretical framework for assimilation of voice. This framework consisted of a set of assumptions. One of the assumptions was that when a syllable is stressed, the (gradual) features of the phonemes belonging to the stressed syllable are "intensified". The implication of this view is that in two-consonant clusters of which the members belong to different syllables of which one is stressed, only the features of the consonant belonging to the stressed syllable are made more prominent. Since before assimilation takes place, the first consonant is voiceless and the second voiced in Dutch, this means that in pre-stress position the effect of stress would lead to the domination of the voiced consonant, and in post-stress position to the domination of the voiceless consonant. As a consequence, we expect that before stress more regressive assimilation will occur than after stress. This expectation was not conclusively confirmed experimentally (Slis 1981 b); only in stop-stop clusters we found significantly more ($p<.05$) regressive assimilation before than after stress. In fricative-stop clusters no significant differences were observed. In the discussion of the results, however, we made the reservation that an uncertain factor in this experiment was that all clusters before stress were situated around a word boundary (external assimilation), while most of the clusters after stress were situated within compound words (internal assimilation).
From a pilot experiment (unpublished) we got the impression that the difference between internal and external assimilation counteracted the effect of position relative to stress. Therefore, we decided to test the hypothesis once more that more regressive assimilation occurs before than after stress. In this conclusive experiment we compared two-consonant clusters before and after stress, both across word boundaries. In order to check the supposed effect of internal versus external influence in the first experiment, we added clusters after stress within compound words, which could be compared with stress across word boundaries. Clusters before stress in compound words were not included since compound words are generally stressed on the first constituent in Dutch.
As we found differences in assimilation between speech of men and women (Slis 1981 c) we used speech samples from 10 men and 10 women.

2. METHOD

Ten male and 10 female subjects, members of the staff of the Institute of Phonetics and linguistics students, were asked to read aloud 21 sentences in which syllables that had to be stressed were underlined. The sentences contained 6 stop-stop (viz. /pb,tb,kb,pd,td,kd/) and 6 fricative-stop clusters (viz. /fb,sb,xb,fd,sd,xd/) in pre- and post-stress position across word boundaries, and in post-stress position within compound words. The experimental method is extensively described in Slis (1981 a). In short, two signals were tape recorded, viz.:

1. the speech signal;
2. the output of an electrolaryngograph.

On the basis of UV-oscillograms of these two signals we scored assimilation in three categories, viz.:

1. no assimilation: voice tail shorter than 50 ms and V.O.T. before opening of the vocal tract;
2. regressive assimilation: voice tail longer than 50 ms and V.O.T. before opening of the vocal tract; this includes continuation of voicing;
3. progressive assimilation: voice tail shorter than 50 ms and V.O.T. at or after the moment of opening the vocal tract.

We also measured the consonant durations between the moments of oral closing and opening from the oscillograms.

3. DATA ANALYSIS

Since we wanted to know whether it was justified to summate the data of men and women and those of stop-stop and fricative-stop clusters, we first performed a multidimensional chi-square test (log. lineair model of Goodman (1971); an extension of the normal chi-square test for more than two independent variables). With this test it is possible to detect interactions between the independent variables, viz. sex of the speaker, cluster type, and stress position & boundaries, with respect to the dependent variable, viz. assimilation of voice.

A three way analysis of the results showed no interaction between sex, stress & boundary and assimilation. However, with regressive assimilation a significant interaction was observed between sex and cluster type. The latter is mainly due to a very low frequency of regressive assimilation with female speakers; only one case of regressive assimilation was found after stress across word boundaries. Nevertheless we concluded that it was justifiable to regard the influence of sex as being independent of the other independent variables. In the further analysis of the results we therefore considered the pooled data of men and women.

Between cluster type, stress & boundary, and assimilation a significant interaction was observed. As a consequence, we decided to analyse stop-stop and fricative-stop clusters separately with a normal chi-square test. In cases where the frequency in one of the cells of the matrix was lower than 5 we used the (equivalent) L-test (Spitz 1968). For the comparison of the mean durations we used the t-test.

4. RESULTS

The results with respect to the frequency of assimilation are summarized in Fig. 1, those with respect to durations in Table I. In Fig. 2 we compared the data of male and female speakers, and of stop-stop and fricative-stop clusters.

5. DISCUSSION

5.1 OCCURRENCE OF ASSIMILATION

5.1.1 Position of stress: In stop-stop clusters assimilation occurred less after than before stress. In a parallel experiment (unpublished) we observed a strong tendency among our subjects to introduce pauses after a stressed syllable. Van Hooff and Van den Broecke (1983) showed that the frequency of assimilation decreases with increasing depth of linguistic boundaries. We think that the introduction of pauses and the decrease of assimilation are both related to the depth of linguistic boundaries.

Although no difference in frequency of assimilation is observed in fricative-stop clusters before and after stress, the tendency to introduce pauses manifests itself in longer cluster durations after than before stress. De Rooij (1979) observed that pauses, syllable lengthening, and pitch inflections contribute to the perception of prosodic boundaries. Coker and Umeda (1973) speak of pseudo-pauses in cases where a pause is suggested while no actual silence occurs.

We assume that both the high incidence of "no assimilation" after stress in stop-stop clusters and the longer duration of clusters with "no assimilation" in fricative-stop clusters in our material are signs of deeper syntactic boundaries after stress than before stress.

5.1.2 Word boundary: A comparison of clusters across word boundaries with clusters in compound words - both after stress - shows that assimilation occurs less frequently across word boundaries than within words in stop-stop clusters and that cluster durations are longer with fricative-stop and stop-stop clusters. These results are fully in line with the results obtained and explained above; the linguistic boundary is less deep within words than across word boundaries and consequently more assimilation in stop-stop clusters and shorter cluster durations in fricative-stop clusters are observed within words.

5.2 DIRECTION OF ASSIMILATION

5.2.1 Position of stress: The results show that considerably more regressive assimilation and less progressive assimilation occurs before than after stress with both cluster types. This confirms our hypothesis; before stress the voiced second consonant exerts a major influence on assimilation, while after stress the voiceless first consonant predominates. Fricative-stop clusters after stress proved to have significant longer durations than before stress. This seems to corroborate a similar difference with not-assimilated fricative-stop clusters signalling the depth of a syntactic boundary (see above).

5.2.2 Word boundary: Post-stress clusters within compound words showed two to three times more regressive assimilation than across word boundaries. This is mainly due to an increase of assimilation within words. This also corroborates the results of Van Hooff & Van den Broecke (1983). Cluster duration within words was significantly shorter than across boundaries under nearly all conditions (regressive, progressive and no assimilation with both cluster types). We take it that this also indicates a stronger coherence of articulation within words.

5.3 SEX OF THE SPEAKER

Under all conditions (before and after stress across word boundaries and after stress within words in stop-stop and fricative-stop clusters) more progressive and less regressive assimilation is observed in male speech than in female speech (significant in 5 of 6 conditions). These results confirm previously obtained data (Slis 1981 c). With men the frequency of regressive assimilation is nearly twice that of progressive assimilation, while, contrarily, with women the frequency of progressive assimilation is twice of regressive assimilation.

5.4 CLUSTER TYPE

Under all conditions more progressive and less regressive assimilation is found in fricative-stop clusters than in stop-stop clusters. More cases of no-assimilation are observed with fricative-stop than with stop-stop clusters (significant in 5 of 6 conditions). The results confirm previously obtained data (Slis 1981 c & 1982). We see that in stop-stop clusters about 1.5 times more regressive than progressive assimilation occurs, and in fricative-stop clusters nearly two times as much progressive as regressive assimilation is found.

The general phonological rule that obstruent-stop clusters show regressive assimilation (Trommelen & Zonneveld 1979) is not confirmed by our results. On the basis of the present results we can at best say that in obstruent-stop clusters no general preference for the direction of assimilation is found.

6. CONCLUSION

Summarizing we found that position of stress and word boundaries played a role in assimilation. In previous studies we studied mainly speech of men in pre-stress position. An important conclusion from these former experiments was that, in obstruent-stop clusters, mainly regressive assimilation occurs. If, however, we include data of post-stress clusters and data obtained with female speakers, we have to change this conclusion: On average, equal frequencies of regressive and progressive assimilation of voice are found in obstruent-stop clusters in Dutch.

7. REFERENCES

Erp, A. van, Halfens, Y.M.A., Hoogeveen, W.N.J.J., Zink, A.H.M., The effect of stress on assimilation of voice in Dutch and English. Proceedings Institute of Phonetics Nijmegen 5, 1981, 75-83

Goodman, L.A., Partitioning of chi square, analysis of marginal contingency tables and estimation of expected frequencies in multidimensional contingency tables. Journal of the American Statistical Association 66, 1971, 339-344

Hoof, C.M.C. van & Van den Broecke, M.P.R., Assimilation of voice in Dutch at three types of linguistic boundaries. Progress Report of the Institute of Phonetics Utrecht 8:1, 1983, 31-40

Slis, I.H., The effect of speaking rate on assimilation of voice. Proceedings Institute of Phonetics Nijmegen 5, 1981 a, 160-176

Slis, I.H., The influence of the position of a stressed syllable on the assimilation of voice. Proceedings Institute of Phonetics Nijmegen 5, 1981 b, 177-187

Slis, I.H., Half a dozen influences on assimilation of voice in Dutch. Proceedings Institute of Phonetics Nijmegen 5, 1981 c, 188-203

Slis, I.H., Reflections on a theoretical frame for the assimilation of voice. Proceedings Institute of Phonetics Nijmegen 5, 1981 d, 204-215

Slis, I.H., Voice quality and assimilation of voice. Proceedings Institute of Phonetics Nijmegen 6, 1982, 53-60

Spitz, J.C., Statistiek voor psychologen, pedagogen, sociologen (3rd ed.). Amsterdam: Noord-Holland Uitg. Mij., 1968, 354-357

Trommelen, M. & Zonneveld, W., Inleiding in de generative fonologie. Muiderberg: Coutinho, 1979

Table I Mean cluster duration (in ms) per assimilation category for (A) stop-stop and (B) fricative-stop clusters; differences between mean durations (in ms) with regard to assimilation category are listed underneath and with regard to experimental condition are listed right of the respective mean durations (only means based on five or more measurements are indicated)

	across word boundaries		within words			
(A) stop-stop	a. before stress	b. after stress	c. after stress	a-b	a-c	b-c
no assimilation	-	176.9	136.7	-	-	40.2**
with assimilation	128.2	134.4	117.2	- 6.2	11.0**	17.2*
no - with ass.	-	42.5**	19.5			
regressive ass.	128.0	125.4	110.8	2.6	17.2**	14.6*
progressive ass.	130.0	137.9	124.7	- 7.9	5.3	13.2**
progr.-regr. ass.	2.0	12.5*	13.9**			
(B) fricative-stop						
no assimilation	174.6	208.3	165.0	-33.7**	9.6	43.3**
with assimilation	141.6	163.8	150.1	-22.2**	- 8.5*	13.7**
no - with ass.	33.0**	44.5**	14.9**			
regressive ass.	142.1	132.9	135.6	9.2	6.5	- 2.7
progressive ass.	141.1	169.9	159.3	-28.8**	-18.2**	10.6
progr.-regr.ass.	- 1.0	37.0**	23.7**			

* p < .05 (t-test)
** p < .01 (t-test)

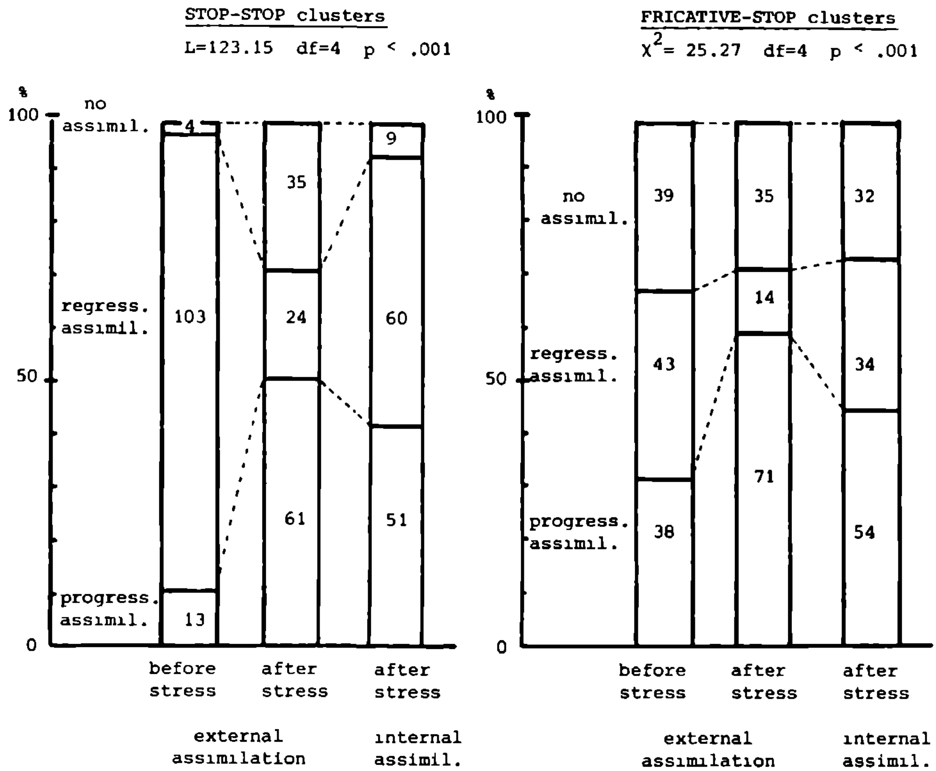


Fig. 1 Frequency of assimilation (in %) as a function of stress and word boundary in stop-stop and fricative-stop clusters (the data of 10 men and 10 women were pooled) with the actual frequencies indicated within the histograms

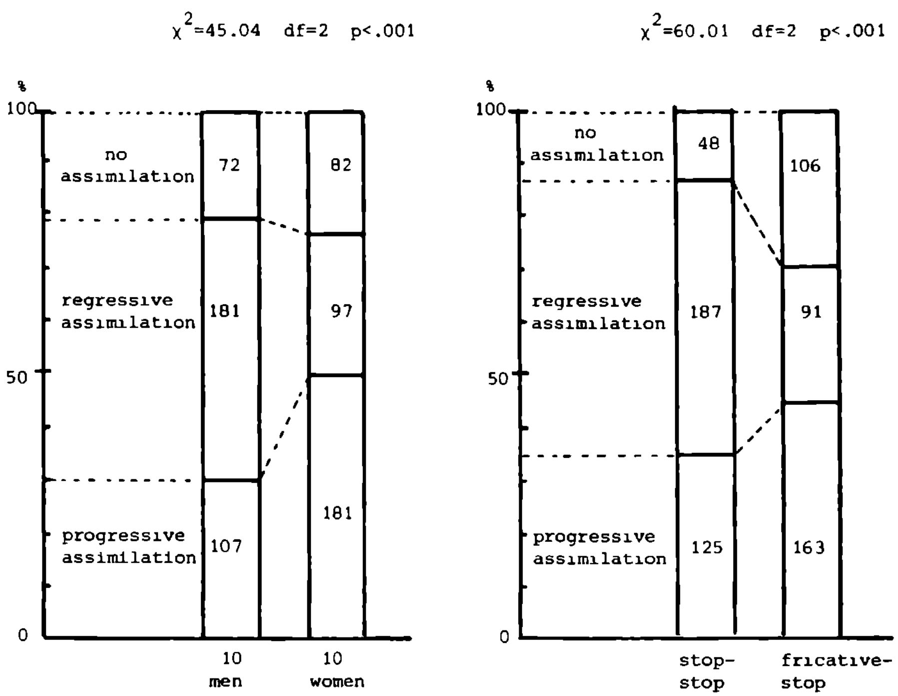


Fig. 2 Frequency of assimilation (in %) as a function of the sex of the speaker (left) and the cluster composition (right), with the actual frequencies indicated within the histograms (data of all stress and boundary conditions are added)

CHAPTER 10

EPILOGUE

10.1 INTRODUCTION

In this epilogue we will first review the results of the measurements on the durational aspects of assimilation, which may be regarded as the manifestation of assimilation of effort.
In the third section the eleven influences on assimilation that were investigated, will be discussed and explained in terms of the five articulatory parameters mentioned in the prologue.
In the intermezzo four different (presumably independent) articulatory features were introduced. In the fourth section we will discuss whether all possible articulatory features are necessary for a phonological description of the voiced-voiceless distinction in Dutch, or whether one phonological feature is sufficient.
In the final section an overall conclusion is given and some remarks will be made on possible future research on the voiced-voiceless distinction.

10.2 TIME PARAMETERS IN ASSIMILATION

In an article not included in this volume (Slis 1975) we paid attention to the timing of oral gestures in two-consonant clusters (C1-C2) in cases where "articulatory effort" was or was not applied in the form of stress on the next syllable. As lengthening of the first consonant (C1) (belonging to the first, unstressed syllable) was observed when the following syllable was stressed, it was concluded that articulatory features belonging to the speech sounds of one syllable can be coarticulated in an adjoining syllable.
In another article (Slis et al. 1977) results of measurements of the duration of a large set of two-consonant clusters with different phonological compositions were reported. The speech material consisted of 22 lists of 10 words each. Each word was repeated 10-20 times within the lists in random order. In all, about 2700 clusters were measured. These results, too, showed that traces of effort which stem from the voicelessness of a consonant can be found in the adjoining consonant; this kind of coarticulation can be regarded as assimilation of effort.

On the basis of the results of the latter measurements seven rules for synthesis by rule were formulated, which are listed below (Slis et al. 1977):
(1) The first rule obtained states that " the moment of closing of voiceless plosives is advanced 15 ms", thus leading to voiceless consonants that are 15 ms longer than voiced ones and to a compensatory shortening of the preceding vowel.
(2) The coarticulation of effort on a phonemic level was demonstrated by the second rule stating that " if the first consonant in a cluster is a nasal and the second consonant a voiceless plosive, the beginning of the nasal is advanced 10 ms". This was interpreted as coarticulation of effort in the (lax) nasal originating from the voiceless plosive. In terms of assimilation we can describe this as regressive assimilation of

effort.
(3) On the other hand the third rule states that " in clusters beginning with a voiceless consonant the advancement of a possible second voiceless plosive is omitted". The measurement data of the heterorganic clusters are however less conclusive than this rule suggests. The mean total duration of /pd/ and /tb/ is about 15 ms shorter than of /pt/ and /tp/. Both C1 and C2 have a shorter closure duration in the voiceless-voiced plosive succession. These data demonstrate that:
(a) if C2 is a voiced plosive, a shorter duration occurs than if C2 is voiceless (this is in accordance with the rule for single intervocalic stops),
(b) regressive assimilation of [- effort] is found which is manifest as a shorter duration of the voiceless C1 before a voiced C2.
The second effect (b) is neglected in the sythesis rules proposed, since the shorter duration of C1 in a stop-stop cluster is not effected acoustically in the total cluster duration; the end of C1 is overlapped by C2 and it will pass unnoticed when the point of overlap shifts in time. Besides, neglecting this effect makes a generalisation of this rule and rule 6 and 7 possible, stating that repetition of a feature of C1 in C2 of a C1C2 cluster results in shortening of C2.
(4) Rule 4: "nasals are lengthened by 10 ms" is not relevant for the voiced-voiceless distinction.
(5) The finding that voiced plosives, which only occur as C2 in Dutch two consonant clusters, are shortened 10 ms, leads to the conclusion that effort of a first voiceless C1 is not progressively assimilated in a voiced C2.
(6) Rule six states that " in clusters with two nasals in succession the second nasal has to be shortened by 20 ms ."
(7) Rule seven states that " homorganic clusters are made 10 ms shorter than heterorganic clusters".

In summary, we may conclude that indications for regressive assimilation of the presence or absence of effort in the second consonant of a two-consonant cluster are found in the average data on which the synthesis rules are based.

The three articles on assimilation of voice (Slis 1983 a, 1983 b, 1985) show that both progressive and regressive assimilation occur in obstruent-stop clusters. We cannot decide whether effort is assimilated progressively or regressively on the basis of average durations, since assimilation of voice may take place in both directions. The best method to decide on the direction of assimilation of effort would have been to use closure duration of the individual consonants in the clusters as a criterion for the presence or absence of effort.
Since no operational tools were available to measure these durations accurately enough, the next best method was chosen. On the basis of voicing, as defined in the introduction (section 1.6) it was decided whether progressive, regressive or no assimilation of voice occurred. Total cluster durations were measured for each cluster. The mean durations of the three assimilation classes were calculated. Clusters with progressive assimilation proved generally to be longer than clusters with regressive assimilation. In Slis (1983 a) these differences lie between -1 and +37 ms. Similar results (viz. differences between -2 and

+69 ms) were obtained with the speech material of the other two articles in this volume; these data are not included in these articles, but can be found in Slis 1981 a and b and Slis 1982 b. These differences proved to be significant (t-test; $p<.01$) in 11 out of 17 comparisons that could be made; the remaining 6 comparisons yielded non-significant differences. The average of the differences is about 15 ms. This corresponds with the difference found between /pt/ and /tp/ vs. /pd/ and /tb/ clusters (Slis et al. 1977). These results support (at least do not contradict) the hypothesis that progressive assimilation of voice is accompanied by a simultaneous progressive assimilation of effort resulting in a longer cluster duration. We assume that this lengthening is the result of a lengthening of the underlying voiced second stop that became [+ tense] and [- voiced].

10.3 JUSTIFICATION OF MEASURED ASSIMILATION IN TERMS OF COARTICULATION

In the prologue (section 1.8) we postulated that assimilation of voice is a special form of coarticulation. On the basis of mechanical and aerodynamic properties of the vocal cords and the vocal tract we defined five parameters which influence vocal cord vibration, viz.:
(A) glottal width,
(B) vocal cord tension,
(C) vibrating mass of the vocal cords,
(D) volume of the supra-glottal cavity,
(E) duration of the oral closure.
These five parameters were held responsible for assimilation of voice. In section 1.7 we mentioned that the condition under which speech is produced, may influence the assimilation observed. We therefore selected eleven speaking conditions which were found to influence assimilation. In this section we will show that elaboration of the five parameters mentioned under A, B, C, D and E led to correct predictions of the experimental results obtained for each of the eleven experimental conditions that were expected to influence assimilation of voice.

1) Monotonous speech needs extra attention to the laryngeal setting compared to spontaneous intonated speech. It was speculated that this extra attention is effected as increased tension of the vocal folds. Although, as far as we know, no physiological evidence is reported regarding this increased tension, it fits with introspective experiences. Extra tension leads to a significant increase of progressive assimilation (Slis 1983 a).

2) A rich literature exists with respect to the relation between pitch and muscular activity; an overview is given by e.g. Sawashima (1974) and Fujimura (1977). Rising pitch is generally accompanied with an increase of laryngeal activity. In singing (e.g. Vennard et al. 1970 a & b) and speaking (e.g. Collier 1975, Atkinson 1978) the cricothyroid muscles stretch the vocal cords, leading to increased tension. Also other intrinsic laryngeal muscles, viz. the lateral cricoarytenoid and the vocalis muscle (Vennard et al. 1970 a & b), show an increase of activity with increasing pitch. The increase in tension of the vocal cords results in a less favourable condition to maintain vocal vibration. As a consequence vocal vibration is interrupted. This

is reflected in an increase of progressive assimilation with rising pitch (Slis 1983 a).

3) Several criteria were used in the classification of good vs poor voices in the experiments described in this volume, viz. voice dynamics in terms of minimal and maximal output sound pressure, pitch range, subjective hoarseness and habitual speaking pitch. Therefore, the poor voice quality of the subjects may be due to various causes. A common factor is a lowered efficiency of voice production; inefficient use of the airstream through the glottis implies that part of the available energy is not used to keep the vocal cords vibrating and that interruption of voicing may result. This can manifest itself as hoarseness (whispery voice) or the inability to vocalise at low airstreams, and of course progressive assimilation. Another characteristic of poor voices is a habitually raised pitch which is caused by too tense vocal cords. We argued above that this also leads to an increase of progressive assimilation. Indeed the results show more progressive assimilation with poor than with good voices (Slis 1983 a).

4) Vocal folds of men are longer and heavier than those of women. These organic differences are the origin of more favourable conditions for maintaining vocal vibration in men than in women. The voiceless interval in intervocalic voiceless plosives is therefore shorter in male than in female speech (Weismer & Fromm 1983). A similar shortening of the voiceless interval in consonant clusters gives rise to an increase of regressive assimilation, since the duration of the voice tails will tend to pass the 50 ms border. Consequently, male speech shows more regressive assimilation than female speech (Slis 1982 a, 1983 a, 1985).

5) At a low speech rate more time is available to realise both consonants, one after the other. With increasing speed the number of clusters in which mutually excluding gestures overlap, increases. As a consequence the number of clusters in which assimilation occurs, is found to increase with tempo. The direction of assimilation, progressive or regressive, is conditioned by the outcome of the battle between the two imcompatible gestures. Therefore we may expect an increase in equal proportions of both progressive and regressive assimilation at the expense of "no assimilation", which is confirmed by the data (Slis 1982 a, 1985).

6) Voiceless fricatives show a larger glottal aperture than voiceless plosives (Slis 1970). It seems therefore probable that more muscular effort for the glottal opening gesture is applied in fricatives than in plosives. Moreover, longer duration is often considered as a sign of more "tension" (Catford 1977). Assuming that peripheral coarticulation is the origin of assimilation, in which stronger commands overrule weaker ones, it must be expected that under circumstances where a voiceless fricative and a voiced plosive overlap, the stronger glottal opening of the fricative outstrips the glottal closing movement of the voiced plosive more frequently than the weaker glottal opening of a voiceless plosive would have done. This hypothesis is confirmed by the data. More progressive assimilation is observed in fricative-stop than in stop-stop clusters (Slis 1982 a, 1985).

On the other hand, voiced fricatives tend to become voiceless in Dutch (Cohen et al. 1971, and Debrock 1977) and voiceless realisations of voiced fricatives are accepted by a majority to be correct Dutch (Gussenhoven 1981 a & b, Gussenhoven & Bremmer 1983). From these indications it can be inferred that the voicing effort in voiced fricatives is low compared to the devoicing effort in voiceless plosives and fricatives. In a competition between a first voiceless fricative or plosive and a successive voiced fricative, the glottal opening activity of the first voiceless obstruent will be stronger than the glottal closing activity of the successive fricative. Consequently the cluster devoices, which becomes evident as progressive assimilation.
In addition to the progressive assimilation due to a mechanical origin, some speakers are inclined to devoice initial fricatives anyway. In this case a voiceless obstruent is followed by a devoiced fricative. This results in a cluster production that satisfies the criteria for progressive assimilation without assimilation actually taking place.
In the results all obstruent-fricative clusters show progressive assimilation (Slis 1982 a, 1985).

7) Voice onset times of plosives are found to be dependent on the place of articulation. The differences are systematically present in all languages investigated (e.g. Lisker and Abramson 1964). Voice onset time proves to be shortest in (bi)labial plosives, and longest in velar plosives; alveolar and dental plosives have intermediate values. Since these differences are not language-specific it seems obvious that we have to look for a common cause. Articulatory restraints are an acceptable candidate. These effects may be caused in the following way:
The oral closing (and opening) gesture in labial plosives is executed by the lips. These articulators can be moved completely independently of the larynx and will consequently have no repercussions for laryngeal functioning.
In alveolar and dental plosives the oral closing (and opening) gesture is executed by the tip of the tongue. Although the tip of the tongue can bend upward without moving the tongue body, we must move the tongue body forward and upward if we want to obtain a complete closure in /t/ and /d/. Analogous to the reasoning by Honda (1983) for high and low vowels, we may assume that the hyoid bone moves forward with the tongue body. Since the superior horns of the thyroid cartilage are connected to the posterior processi of the hyoid bone by the lateral thyroid-hyoid ligaments, the hyoid bone pulls the upper part of the thyroid cartilage forward. The inferior horns of the thyroid cartilage articulate on a fixed point on the sides of the cricoid cartilage. As a consequence, the thyroid cartilage is tilted forward. The anterior end of the vocal folds, which is attached to the thyroid cartilage, moves downward and forward thus stretching the vocal folds. The vibration condition of the vocal folds becomes therefore less favourable and voice onset may be delayed in dental plosives compared to labial plosives. This results also in more progressive assimilation in clusters with /d/ than with /b/ (Slis 1982 a, 1985).
The oral closing (and opening) in velar plosives is performed by the dorsal part of the tongue body. It may be expected that the displacement of the hyoid bone will be larger than with dentals, since the oral closure at a velar place of articulation has to be effected by the tongue

body only. Besides, upward stretching of the surface layers of the pharynx walls (and consequently of the inner walls of the larynx) yields an even more unfavourable vibration condition of the vocal cords. This will result in longer VOT's in velar than in dental and labial plosives, as observed (e.g. Lisker and Abramson 1964). Since in Dutch velar /g/ does not exist, we could obtain no data of /obstruent+g/ clusters.

8) In stressed syllables generally more effort is applied than in unstressed ones. As a consequence the sound level and the pitch are raised, a longer duration occurs, and the articulation is executed more accurately. If we assume that stress is attached to syllables, we may expect that in instances where two syllables, one stressed and one unstressed, meet, the gestures of the stressed syllable will overrule those of the unstressed one. In the concrete case of a succession of a voiceless and a voiced obstruent, this means that the voice aspect of the obstruent belonging to the stressed syllable will dominate over that of the unstressed syllable. A similar transfer of effort was observed with the measurements of the duration in two-consonant clusters (see above, Slis 1972, Slis et al. 1977); duration of a postvocalic C1 was increased by an advancement of the moment of closing the oral tract if the following syllable was stressed. Therefore progressive assimilation is more frequent if the preceding syllable is stressed and the effort level of the voiceless C1 is raised, than if the following syllable is stressed and the effort level of the voiced C2 is raised (Slis 1983 b).

9) In Slis (1971 a) we tried to justify that effort of articulation is effectuated as an advancement of the gestures involved, as in voiceless plosives or stressed syllables, and reversely, if a comparable advancement of an oral gesture is observed, that this might be caused by applying more effort. The argument for reversion is better founded if the extra effort is in accordance with introspective notions. This is the case in the difference between long and short vowels. In German the terms "scharf geschnitten" (sharp cut) and "weich geschnitten" (weak cut) are used for short and long vowels respectively. Sharp cut suggests more effort in the vowel termination than weak cut. On these grounds it can be assumed that the advanced closing after short vowels is due to an increase of articulatory effort of the closing movement, which belongs to the following consonant. As a consequence, the following consonant should be realized with more effort after short vowels. According to the model for the voiced-voiceless distinction, this consonant should show an increase of voiceless elements. This is confirmed by our measurements (Slis 1984); after short vowels a tendency towards more progressive assimilation than after long ones is found.

10) On intuitive grounds it may be expected that the articulatory program of two (or more) subsequent syllables is more coherent if these syllables belong to one than to several words. In this point of view we preserve the possibility that the articulatory program for a word is learned and stored as a whole, and that as a result assimilation within words can not be regarded as coarticulatory effects, but must be due to programmed articulations. Our results are in accordance with this assumption; more assimilation is observed within words than on word boundaries, suggesting more coherence within words. However, a different

distribution of regressive and progressive assimilation is found within words and across word boundaries, suggesting another mechanism (Slis 1983 b, van Hooff & van den Broecke 1983, Loots 1983).
The higher frequency of regressive assimilation within words compared to across word boundaries may be explained by the observation that more regressive assimilation is subjectively perceived than objectively measured (unpublished data). In a listening experiment with part of the material that has been measured for the present experiments on stress, we observed that a number of the clusters with progressive assimilation were perceived as not assimilated and clusters that were not assimilated were perceived to have regressive assimilation. Similar effects can be shown with intervocalic obstruents; obstruents with interrupted voicing can be perceived as voiced if e.g. the duration is very short or if extensive formant transitions are present in the adjoining vowels (Slis and Cohen 1969 a). It seems plausible that if a person learns a new word in which he perceives an assimilated cluster, he may fail to detect that this cluster was originally built up of a voiceless and a voiced consonant. Consequently he stores a word image which includes assimilation. In reproduction he will use an articulatory program with preprogrammed assimilation. In clusters on word boundaries the speaker does not have a preprogrammed assimilation command; a complete new combination will have to be realised.

11) In the Dutch literature on assimilation, differences between the results obtained by different investigators were often attributed to differences in the dialectal background of the investigators and their subjects. Since the University of Nijmegen is a regional university, attended by students from various dialectal regions, we had the opportunity to check whether the dialectal background of the speakers is reflected in the treatment of assimilation in their pronunciation of standard Dutch. As this research does not aim at a description of dialectal influences, and we were not in a position to select speakers from neatly defined dialectal areas, the results obtained were grouped with respect to rather large regions of origin of the speakers. The regions chosen were mentioned in the literature to be relevant for assimilation of voice. In the material presented no significant influences of dialectal background were detected (Slis 1982 a, 1985).

In the above elaboration of the eleven influences studied we observed that it is, in principle, possible to explain the results as a consequence of coarticulation of articulatory features. The relevant features were found in the research on intervocalic voiced and voiceless obstruents and in the literature on articulation in general.

10.4 ONE OR TWO FEATURES ?

A point that needs further discussion concerns the question whether more than one phonological feature is necessary to describe the voiced-voiceless opposition in Dutch. The discussion of the articulatory data in the intermezzo shows that different mechanisms participate in the voiced-voiceless distinction. In languages with more than two functionally different realizations along the voicing dimension, like e.g. Thai, Korean or Hindi, it is evident that more than one binary feature is necessary to obtain a description of these languages. In languages with only two different realizations along the voicing dimension , nobody will disagree that only one feature is sufficient for a purely functional description.

Phonology, however, claims to give more than a purely functional description. According to Crystal (1980:269) "...phonology is concerned with the range and function of sounds in specific languages (and often therefore referred to as ´functional phonetics´), and with the rules which can be written to show the types of phonetic relationships that relate and contrast words and other linguistic units".

Although this definition is restricted to oppositions within one language it seems attractive to enlarge a system describing one language with a few features to obtain better "phonetic relationships". This was done by Fischer-Jørgensen (1969, 1980) with respect to the voiced-voiceless opposition. She proposed three different features to describe the voiced-voiceless opposition, viz. voice, tenseness and aspiration. Her argumentation is based on the fact that the classes indicated by /b,d,g/ and /p,t,k/ are both realized as voiceless plosives in Danish, and that contrary to other languages, the production of /b,d,g/ is accompanied by phenomena that can be interpreted to be [+ tense]. In fact her arguments are based on data borrowed from other languages. An alternative solution to her problem could have been to regard the features "voice" and "tenseness" to be redundant for Danish and to describe the opposition by plus or minus aspirated, as Petursson (1976) did for Icelandic.

Lisker and Abramson (1964) showed that the phoneme boundaries in various languages in which a binary opposition is found between two voicing categories differ with respect to voice onset time (VOT). VOT´s in Spanish, Dutch, English and Cantonese showed a later voice onset in that order for voiced as well as voiceless plosives. It seems likely that more differences in VOT´s would have been observed if data on more languages had been available. Lisker and Abramson (1971) commented that "Our data ... did not provide strong support for rejecting the possibility that speakers are capable of producing stops with any duration of voicing lead or lag over a range of several hundred milliseconds". If, in addition, also other articulatory and/or acoustic phenomena were compared in detail between various languages, it may be expected that the realization of the voiced-voiceless distinction will be found to be different in almost all languages. Those differences are often referred to by the term "bases of articulation" (O´Connor 1973). The restriction "in specific languages" in the definition of Crystal implies that phonologists are relieved of the obligation to create a universal phonological system that covers all different languages.

If, however, one wants to describe all different shades of the voiced-voiceless distinction in all languages, a large system of features will be necessary. Four possible different articulatory mechanisms (which can operate independently of each other) have already been mentioned in the intermezzo. These can be attributed to two groups of phenomena, viz.

(a) supra-glottal events such as oral closure duration, degree of EMG-activity and volume of the supra-glottal cavity, and

(b) glottal events, viz. abduction vs adduction of the vocal folds by means of different patterns of muscle activity, and the timing of these gestures.

The first group refers to the opposition tense-lax and pharyngeal expansion, the second to the voiced-voiceless distinction and aspiration. More articulatory features can be added; for instance, Halle and Stevens (1971) proposed a system with four laryngeal features, viz. stiff vocal cords, slack vocal cords, wide glottis and narrow glottis. In doing so they added stiffness of the vocal cords as an independent articulatory mechanism to glottal aperture.

The timing of the various articulatory features with respect to each other may also be added as separate features, thus yielding a universal system with which it is presumably possible to describe all nuances in all languages. This system, however, will be very complicated and un-economical. In our case it is wise not to pay attention to differences between languages. On the basis of the data presented for single consonants we conclude that one phonological feature is sufficient in languages like Dutch, in which a binary distinction along the voicing dimension is functional. This phonological feature may be built-up of a bundle of articulatory features that are language specific.

In the case of assimilation of voice, however, several authors indicate the necessity for separate "voiced-voiceless" and "tense-lax" features (e.g. Booij 1981, Hubers and Kooij 1973, Trommelen and Zonneveld 1979). The question is dealt with in Slis (1982 a, 1985). The phonological rules suggested predict progressive assimilation of effort and voice in obstruent-fricative clusters. In these clusters effort and voice are rule governed in exactly the same way; since the separate rules are redundant in this case, effort and voicing can be regarded as one feature.

For obstruent-stop clusters these phonological rules predict regressive assimilation of voice and progressive assimilation of effort which yields clusters that are built up of two consonants that are both [+ voiced] and [+ tense].

A first objection against this phonological description is that our measurements show longer cluster durations ([+ tense]) in clusters with progressive assimilation of voice ([- voice]) than in clusters with regressive assimilation of voice ([+ voice]).

A second objection is based on calculations with articulatory models. The articulatory consequence of a [+ tense] articulation is a relatively small supra-glottal cavity with stiff walls. The outcome of calculations with the model of Rothenberg as presented by Muller (1983) is that voicing must be expected to cease within 50 ms under [+ tense] conditions (see intermezzo, paragraph 6.3). Voicing will be resumed at the opening of the vocal tract. In our terms we categorize a similar [+ tense] realisation of the cluster as progressively assimilated, since the voice tail is shorter than 50 ms and VOT is 0 ms. This is in conflict with the prediction of the rule.

Apart from the first two objections, the [+ voiced, + tense] description does not differentiate between conditions under which regressive and progressive assimilation occur. They only predict that no "pure" voiced or voiceless cluster will be realized. From the measurements we learned that rough predictions of the direction of assimilation must be possible if only enough features are introduced,

like sex, emotion, voice quality etc.
On the other hand, if only one feature is used in phonology, predicting progressive assimilation in obstruent-fricative clusters and no assimilation in obstruent-stop clusters, we can explain the observed assimilation patterns by coarticulation on the phonetic level, as argued above. In addition, regressive assimilation in obstruent-stop clusters cannot change the meaning of the word the assimilated consonant belongs to, because the voiced-voiceless distinction is redundant for final consonants. Progressive assimilation in obstruent-stop clusters may change the meaning of the word of which the assimilated fricative is the initial consonant; assimilation in the latter clusters can be described with one feature only. We may conclude therefore that two separate features are not neccessary for a phonological description.
The only reason left to introduce two features must be looked for on a phonetic level, viz. in order to obtain a description that takes care of context dependent variants of phonemes (allophones). Here, we would have to prove that events that accompany the voicing distinction occur independently from those that accompany the effort distinction. A longer duration is taken to be a realisation of more effort ([+ tense]) and vocal vibration a realisation of voicedness ([+ voice]). If in obstruent-stop clusters both consonants are [+ voiced, + tense], about equal durations in progressively and regressively realized clusters must be expected, since the latter difference is due to coarticulation only. Contrary to this expectation, the data show longer durations in obstruent-stop clusters with progressive assimilation than with regressive assimilation on several occasions (Slis 1981 a and b, 1982 b, 1983 a). Therefore it may be concluded that [+ voice] is coupled to [- tense] and vice versa. Consequently no need is found to separate the voice and effort features.

10.5 POSSIBLE FUTURE PLANS FOR RESEARCH ON VOICING

A general idea about the articulation of speech sounds that gradually took shape after the study of the voiced-voiceless opposition, is that each speech sound is the result of an articulatory program which is built up of commands to more or less independent articulatory mechanisms. Each resulting articulatory gesture (here called articulatory feature) influences one or more articulatory parameters, e.g. the opening gesture of the glottis may have consequences for the vocal cord tension, the vibrating mass of the vocal folds, the glottal aperture and the glottal resistance for the expiratory airstream. The relative timing of the commands and the degree of innervation are part of the articulatory program.
The suggestion made in this volume is that one phonological feature may be built up of a number of articulatory features. The number of articulatory features that can be taken together to form one phonological feature must be dependent on the function of the feature in the language, and is therefore language dependent. In cases where a combination of articulatory features always occurs in the same composition in a phonemic opposition, it is suggested that for the sake of economy this combination is regarded as one phonological feature only.
The results presented indicate that the articulatory features [voice], [tenseness] and [pharyngeal expansion (or constriction)] always occur in

the same combination in Dutch. Even the phenomena observed in assimilation can be regarded as originating from peripheral coarticulation of two invariant bundles of articulatory features. Whether this feature is called [voice] or [tenseness] is not essential; on different occasions different aspects may be prominent. Similar cases are found in other phonological oppositions; e.g. the tense-lax opposition with vowels is often referred to as long-short. Again, the different terminologies indicate different aspects of the same opposition.

Several assumptions are made in the interpretation of the results, among others with respect to the conditions under which the vocal cords remain in vibration and under which the pulsed air-flow may give rise to a perceivable acoustic output. Different research groups are developing computer models of the vocal cords. For instance a two-mass model of the vocal cords which is coupled to a vocal tract is being built in the Institute of Phonetics in Nijmegen (IFN) with the object of simulating the behaviour of the vocal cords during the production of voiced and voiceless plosives (Cranen 1983).
Another line of future research lies in the field of speech synthesis by rule. The results of the earlier investigations on the voiced-voiceless distinction are implemented in the speech synthesis system described by Slis (1978). Rules for assimilation of voice still have to be formulated for Dutch. A first step has been performed by the experiments described. Perceptual research is also needed to obtain acceptable rules. Preliminary experiments (unpublished material) showed that the relation between the timing of voice activity, which formed the basis for the operational definition of assimilation used in this volume, and the perception of the voicing of the obstruents in a two-consonant cluster (in fact perception of assimilation) is not simple. Various acoustic parameters will have to be tested as to their perceptual contribution to the perception of assimilation in a similar way as with single stops and fricatives. This can be done by making use of synthetic speech. The situation is more complicated than with single obstruents since it has been shown that variations in the acoustic parameters of one consonant in a cluster may have consequences for the perception of the voice character of the other (Wingate 1982). This line of research is also part of the research program of the IFN.
Applications of the results can be expected in voice pathology. In assimilation of voice, the boundary conditions between cessation or continuation of voice activity are reached. Our results showed that more progressive assimilation occurs in poor than in good voices. More research is needed in order to develop a diagnostic tool (instrumental or perceptual) on the basis of assimilation. Further research is also needed to learn whether the assimilation exercises used in speech therapy improve the voice or only suggest a better voice because more regressive assimilation is perceivable.
Gussenhoven and Broeders (1976) point out that speakers are often not aware of the assimilations they make in their mother tongue. If a student is not aware of the assimilations in his own language, it must be difficult to teach him different assimilations in the second language. Persistent errors in the second language may be the result. More comparative research of assimilation is therefore necessary for

applications in foreign language teaching, such as done by Van Dommelen (1983).
In addition to the research that focuses directly on the production or perception of assimilation, future research is to be expected in which assimilation plays a role. For instance, words may obtain a different meaning as a result of assimilation. This effect may play a role in word recognition. In the phonetics department of the University of Leiden experiments have been started that are based on this and similar effects. Indications are reported that the occurrence of assimilation between two consonants that form a cluster depends on the depth of syntactic boundaries (Van Hooff and Van den Broecke 1983). Future research is necessary to find out whether the perception of syntactic boundaries is influenced by assimilation. A similar influence might be useful to mark syntactic boundaries in synthetic speech.
Also in the field of socio-linguistics and dialectology assimilation of voice can be expected to be an experimental variable. Gussenhoven (1981 a and b) showed that the voicing of fricatives is a socio-linguistic marker. It seems probable that a preference for progressive or regressive assimilation is dependent on the socio-economic status of the speakers.

As for the methodological approach for future research we may conclude from the examples given that here a wide variety of possibilities exists. Complicated quantitative models need intricate, up-to-date computer techniques. On the other hand, perceptual experiments on single acoustic parameters can be performed with standard speech synthesizers, or even simple tape-splicing techniques. For socio-linguistic and dialectal research complex statistical analyses have to be performed.

In short we may conclude that research on the voiced-voiceless opposition and on assimilation of voice is far from complete.

SUMMARY

THE VOICED-VOICELESS DISTINCTION AND ASSIMILATION OF VOICE IN DUTCH

The difference between voiced and voiceless obstruents has of old been a topic among linguists and phoneticians; already Petrus Montanus paid attention to the subject in the 17-th century. In the 1940's new aids for speech research became available as a result of an explosive development of electronics. This formed a powerful stimulation for phonetic research. As a result the knowledge about variables that play a role in the voiced-voiceless distinction rapidly increased. Research during the last decades can be divided into two mainstreams, viz.:
1) A stream of research mainly aiming at temporal aspects of speech. For this type of research it is necessary to segment speech. Therefore, segmental features are a central issue in the temporal approach.
2) A stream of research mainly aiming at the frequency domain. In this type of research, segmental aspects are less prominent. Emphasis lies on composition of, and changes in, the acoustic signal. The research into the role of formant transitions is an example.
The experiments reported in this dissertation are inspired to a large degree by problems experienced in the synthesis of speech by means of a synthesizer based on a segmental approach. An important feature of the speech synthesizer was the control of the temporal build-up.

The considerations summarized above are elaborated in the first chapter (prologue). The experiments performed are reported in articles, seven of which have been selected for this volume. They illustrate the development of our research. The first four articles deal with experiments on intervocal voiced and voiceless obstruents (VCV). In the remaining three articles research was extended to combinations of two obstruents belonging to different syllables (VC#CV); Throughout this volume these combinations will be called clusters.

Where voiceless and voiced speech sounds meet (and partly overlap) mutual influencing occurs. In cases where the voice character of one obstruent adapts itself to that of the other one speaks of assimilation of voice. Assimilation may occur because of two mechanisms, viz.:
1) In a quick succession of a voiceless and voiced obstruent, the speaker adjusts the articulation program for voicing in such a way that either both become voiceless (progressive assimilation) or both become voiced (regressive assimilation). To this end a speaker needs internalized rules which prescribe when progressive and when regressive assimilation is to be used.
2) Assimilation results automatically since two conflicting commands are given in quick succession to one articulator. Because of the inertia of the system aspects of both speech sounds will overlap. The ultimate result will depend on both mechanical and aerodynamic properties of the articulatory mechanism; such influencing of successive speech sounds on each other is called coarticulation.
In the introduction we apt for the second hypothesis, which form the

base of the discussion of assimilation.

Two articles (Slis and Cohen 1969 a and b), given in the second and third chapter, reflect the state of the art in the initial phase of our research on the voiced-voiceless distinction, viz. the 1960's. In the first article an inventory of the various aspects of this distinction is made; the second article is an attempt to relate these aspects to each other.
A third article (Slis 1970), given in the fourth chapter, is devoted to a series of additional measurements dealing with the temporal structure of words of which voiced or voiceless obstruents form a part. A model is formulated based upon the results of these measurements and the data froms the first two articles. In this model we tried to attribute all aspects of the voiced-voiceless distinction to one single cause.
In a fourth article (Slis 1971, chapter five) we proceeded along this line of research. With the aid of a number of electromyographic measurements and data on articulation found in the literature we argued that the differences between voiced and voiceless obstruents, long and short vowels, and stressed and unstressed speech sounds can all be attributed to the same articulatory feature, viz. articulatory effort.

The attempt to attribute the large number of aspects that accompany the voiced-voiceless distinction to one single cause, was inspired mainly by the fact that this distinction can be described by one single phonological feature. This attempt seemed succesful around 1970. More recent experimental results forced us to revise the proposed solution. The model of chapter five is critically reviewed and extended in a sixth chapter (intermezzo). It was no longer tenable to proceed from one hypothetical articulatory cause, but we had to assume at least two (and probably more) independently programmable causes. A large number of the aspects could be dealt with by means of the original "effort"-model; with respect to glottal behaviour we had to assume an independent control of the larynx. Moreover, there are arguments to assume an independent control of the volume of the mouth-throat cavity and of the timing of glottal behaviour.

After the discussion of single, intervocal obstruents in the first part of this dissertation we turn to the issue whether the features of the voiced-voiceless distinction manifest themselves in the same way in two-consonant clusters.
The influence of various experimental conditions on assimilation of voice is discussed in chapters seven, eight and nine. Chapter seven (fifth article, Slis 1983 a) deals with the influence of voice quality ("logopedic good" vs. "poor" voices) and pitch (intonated vs. monotonous speech, and monotonous speech with a low, medial and high pitch). The variables discussed in chapter eight (sixth article, Slis 1985) are: phonological composition of the cluster (stop-stop, fricative-stop, stop-fricative, and fricative-friactive), rate of speech (slow, normal, fast), sex of the speaker, and region of origin of the speaker. In the ninth chapter (seventh article, Slis 1983 b) attention is paid to the place of stress in relation to the cluster and to the fact whether a cluster bridges a word boundary (external assimilation) or a word-internal syllable boundary (internal

assimilation).

The purpose of the tenth and final chapter (epilogue) is threefold. In the first part the results of the experiments regarding clusters are discussed. By means of references to other articles (which are not included in this dissertation) it is illustrated in which way the results are used in rules for speech synthesis. Further, it is shown that the results of the experiments on assimilation can all be interpreted in terms of coarticulation. It is argued that the most important parameters can all be viewed as aspects of properties of the vocal cords.
The second subject in the epilogue is mainly of phonological relevance. On the basis of the phonetic data gathered in the course of our investigations it is argued that one phonological feature is sufficient for the description of the voiced-voiceless distinction; if coarticulation is held responsible for assimilation (progressive or regressive), we see no reason to assume more than one distinctive feature.
In a concluding section we speculate on possible future plans for research on assimilation. On the one hand a number of fields in which the results might be applicated are indicated, on the other hand, it is argued that "assimilation of voice" as such has a large number of aspects that call for future fundamental research.

SAMENVATTING

HET STEMHEBBEND-STEMLOOS ONDERSCHEID EN ASSIMILATIE VAN STEM IN HET NEDERLANDS

Het verschil tussen stemhebbende en stemloze obstruenten is vanouds een onderwerp waar taalkundigen en fonetici zich mee bezig hebben gehouden; Petrus Montanus besteedde hier al in de 17de eeuw ruime aandacht aan. Omstreeks de 40tiger jaren van deze eeuw kwamen door een explosieve ontwikkeling van de electronica nieuwe hulpmiddelen voor het spraakonderzoek beschikbaar. Dit was een krachtige stimulans voor het fonetisch onderzoek, waardoor ook de kennis over de factoren die in het stemhebbend-stemloosverschil een rol spelen, snel toe nam. Het onderzoek in de laatste decennia is in twee stromen onder te verdelen, t.w.:

1) Een stroom die zich vooral richt op de temporele aspecten van spraak. Voor dit type onderzoek is het noodzakelijk dat de spraak wordt verdeeld in segmenten. Bij het temporele onderzoek stonden daarom vooral segmentele eigenschappen centraal.

2) Een stroom die zich vooral richt op aspecten uit het frequentiedomein. Hierbij treden segmentele eigenschappen minder op de voorgrond. De nadruk ligt in deze stroom meer op samenstelling van en veranderingen in het akoestisch signaal. Het onderzoek naar de rol van formanttransities kan hierbij als voorbeeld dienen.

Het onderzoek waarvan in dit proefschrift verslag wordt afgelegd, is in belangrijke mate geinspireerd door problemen die werden ondervonden bij spraaksynthese met behulp van een op een segmentele benadering gebaseerde synthetisator. Een belangrijk aspect van de gebruikte spraaksynthese was de temporele structuur.

In het eerste hoofdstuk (proloog) zijn de hierboven samengevatte overwegingen uitgewerkt. Over het verrichte onderzoek is in de vorm van artikelen gepubliceerd. Hieruit zijn zeven artikelen gekozen die bij elkaar een overzicht geven van de ontwikkeling van het onderzoek. De eerste vier hiervan behandelen experimenten over stemloze en stemhebbende obstruenten die ingebed waren tussen twee klinkers (VCV). In de overige drie artikelen werd het onderzoek uitgebreid tot combinaties van twee obstruenten die tot verschillende lettergrepen behoren (VC#CV); dergelijke combinaties worden in dit proefschrift aangeduid met clusters.

Op plaatsen waar stemloze en stemhebbende klanken in spraak op elkaar volgen (en waar ze elkaar deels overlappen) treedt wederzijdse beinvloeding op. Indien daarbij het stemkarakter van de ene obstruent zich (min of meer) aanpast aan dat van de andere, spreekt men van assimilatie. Het optreden van assimilatie kan op twee wijzen tot stand komen:

1) De spreker past zijn articulatieprogramma voor de spieren die bij de stemgeving betrokken zijn zodanig aan, dat bij een snelle opeenvolging van een stemloze en een stemhebbende obstruent, beide stemloos (progressieve assimilatie) of beide stemhebbend (regressieve assimilatie)

worden. Hiervoor zal de spreker over geïnternaliseerde regels moeten beschikken die hem voorschrijven wanneer hij progressieve dan wel regressieve assimilatie moet gebruiken.
2) Assimilatie komt automatisch tot stand doordat zeer snel na elkaar conflicterende opdrachten aan een articulator worden gegeven. Door de traagheid van het systeem zullen aspecten van de beide klanken overlappen. Het uiteindelijk resultaat zal daarbij afhangen van de mechanische en aerodynamische eigenschappen van het articulatiemechanisme; een dergelijke beinvloeding van elkaar door opeenvolgende spraakklanken noemt men ook coarticulatie.
In de inleiding spreken wij ons uit voor de tweede hypothese, die als basis dient voor de verdere behandeling van assimilatie.

In het tweede en derde hoofdstuk volgen twee artikelen (Slis en Cohen 1969 a en b) die de stand van zaken ten tijde van de beginfase van het onderzoek met betrekking tot het stemhebbend-stemloos onderscheid weergeven, t.w. de zestiger jaren. In het eerste artikel worden voornamelijk de verschillende aspecten van het onderscheid geinventariseerd, terwijl in het tweede artikel een poging wordt gedaan deze aspecten aan elkaar te relateren.
In het vierde hoofdstuk volgt een derde artikel (Slis 1970) waarin een serie aanvullende metingen wordt behandeld op het gebied van de tijdstructuur van woorden waarin stemhebbende of stemloze obstruenten voorkomen. Op basis van de gegevens uit deze metingen en de gegevens uit de eerste twee artikelen wordt een model voorgesteld waarmee gepoogd wordt om alle aspecten van het verschil tot een enkele articulatorische oorzaak terug te brengen.
In het vijfde hoofdstuk (vierde artikel, Slis 1971) wordt deze weg verder gevolgd. Aan de hand van een aantal electromyografische metingen en literatuurgegevens over articulatie wordt aannemelijk gemaakt dat het verschil tussen stemhebbende en stemloze obstruenten, tussen lange en korte klinkers en tussen beklemtoond en onbeklemtoond uitgesproken klanken alle teruggebracht zouden kunnen worden tot een en hetzelfde articulatorische kenmerk, namelijk articulatorische inspanning ("effort").

De pogingen om het grote aantal aspecten die het verschil tussen stemhebbende en stemloze obstruenten begeleiden terug te brengen tot een enkele oorzaak werden vooral ingegeven door het feit dat dit verschil zich door een fonologisch kenmerk liet beschrijven. Dit bleek omstreeks 1970 redelijk goed mogelijk. Na die tijd zijn er echter nieuwe onderzoeksgegevens beschikbaar gekomen die het nodig maken om de voorgestelde oplossing te herzien. In een zesde hoofdstuk (intermezzo) wordt het in het vijfde hoofdstuk voorgestelde model kritisch besproken en aangevuld. Er kon niet meer van een oorzaak uitgegaan worden, maar er moesten minstens twee (en mogelijk meer) onafhankelijk van elkaar te programmeren oorzaken verondersteld worden. Een groot aantal van de verschijnselen konden met het oude "effort"-model afdoende beschreven worden; voor het glottale gedrag moest echter een aparte besturing van de larynx verondersteld worden, terwijl er bovendien argumenten zijn die voor een onafhankelijke beregeling van het volume van de mond-keelholte en van de timing van het glottale gedrag pleiten.

Nadat de enkele, intervocale obstruenten in de eerste helft van dit proefschrift aan de orde zijn geweest, wordt in het tweede deel verslag gedaan van het onderzoek waarin werd getoetst of de kenmerken van het stemhebbend-stemloosverschil zich op de zelfde wijze in twee-consonantclusters manifesteren.
In het zevende, achtste en negende hoofdstuk wordt de invloed van verschillende experimentele condities op assimilatie van stem onderzocht. In het zevende hoofdstuk (vijfde artikel, Slis 1983 a) wordt de invloed van de stemkwaliteit ("logopedisch goede" vs. "slechte" stemmen) van de spreker en toonhoogteaspecten (geintoneerde vs. monotone spraak, en monotone spraak met lage, gemiddelde en hoge toonhoogte) besproken. In het achtste hoofdstuk (zesde artikel, Slis 1985) zijn de variabele condities: fonologische samenstelling van de cluster (plofklank-plofklank, plofklank-fricatief, fricatief-plofklank en fricatief-fricatief), spreektempo (langzaam, normaal, snel), geslacht van de spreker en streek van herkomst van de spreker. Het negende hoofdstuk (zevende artikel, Slis 1983 b) handelt over de invloed van de plaats van klemtoon t.o.v. de cluster en het feit of de cluster een woordgrens of een (woordinterne) lettergreepgrens overbrugt.

Het tiende en laatste hoofdstuk (epiloog) heeft een drieledig doel. In de eerste helft worden de resultaten van de experimenten met clusters besproken. Aan de hand van verwijzingen naar andere (niet in dit proefschrift opgenomen) artikelen wordt aangegeven hoe de resultaten in regels voor de spraaksynthese zijn verwerkt. Daarnaast wordt aangetoond dat de resultaten van de assimilatie-experimenten alle geinterpreteerd kunnen worden op basis van coarticulatie. De belangrijkste parameters hierbij zijn terug te voeren op eigenschappen van de stembanden.
Het tweede onderwerp in de epiloog is van fonologisch belang. Op grond van de fonetische gegevens die met dit onderzoek zijn verzameld, wordt beargumenteerd dat voor het stemhebbend-stemloosonderscheid met een enkel kenmerk kan worden volstaan; indien coarticulatie de wijze van assimilatie (progressief of regressief) bepaalt, is er ook binnen clusters geen reden om meer dan een distinctief kenmerk aan te nemen.
In een afsluitende paragraaf wordt tenslotte gespeculeerd over de toekomstperspectieven van het onderzoek aan assimilatie. Enerzijds worden enkele mogelijke practische toepassingsgebieden aangegeven, anderzijds wordt betoogd dat het onderwerp "assimilatie van stem" ook in de toekomst veel aspecten heeft die nog verder onderzoek noodzakelijk maken.

REFERENCES

Abramson, A. S. Laryngeal timing in consonant distinctions. Haskins Laboratory Status Report on Speech Research, SR-47, 1976, 105-112. Also, paper read at the 8th International Congress of Phonetic Sciences

Atkinson, J. E. Correlation analysis of the physiological factors controlling fundamental voice frequency. Journal of the Acoustical Society of America, 63, 1978, 211-222

Booij, G. E. Generatieve fonologie van het Nederlands. Utrecht/Antwerpen: Uitg. Het Spectrum, 1981

Catford, J. C. Fundamental Problems in Phonetics. Edinburgh: Edinburgh University Press, 1977

Cohen, A. Phonetic research. Philips Technical Review, 25, 1964,43-48

Cohen, A., Schouten, j. F., `t Hart, J. Contribution of the time parameter to the perception of speech. Proceedings of the 4th International Congress of Phonetic Sciences, Helsinki. Den Haag: Mouton, 1961, 555-560

Cohen, A., Slis, I. H., `t Hart, J. Perceptual tolerances of isolated Dutch vowels. Phonetica, 9, 1963, 65-78

Cohen, A., Ebeling, C. L., Fokkema, K., Holk, A. G. F. van Fonologie van het Nederlands en het Fries (2de druk, 5de oplage). ´s Gravenhage: Martinus Nijhoff,1971

Cohen-Stuart, A. B. F,S-V,Z Eene bijdrage tot de Nederlandsche uitspraakleer. de Taal en Letterbode, III, 1872, 298-305

Cole, R. A. & Scott, B. Toward a theory of speech perception. Psychological Review, 74, 1967, 431-461

Collier, R. Physiological correlates of intonation patterns. Journal of the Acoustical Society of America, 58, 1975, 249-255

Collier, R., Lisker, L., Hirose, H., Ushijima, T. Voicing in intervocalic stops and fricatives in Dutch. Journal of Phonetics, 7, 1979, 357-373

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M., Gerstman, L. J. Some experiments on the perception of speech sounds. Journal of the Acoustical Society of America, 24, 1952, 597-606

Cranen, B. Preliminary results on the simultaneous registrations of seven laryngeal parameters. Proceedings of the Institute of Phonetics Nijmegen, 7, 1983, 31-45

Cranen, B., & Boves, L. Calibrated recordings of sub-, supra-, and transglottal pressure. Proceedings of the Institute of Phonetics Nijmegen, 8, 1984, 30-47

Crystal, D. A First Dictionary of Linguistics and Phonetics. London: Andre Deutsch Ltd, 1980

Debrock, M. An acoustic correlate of the force of articulation. Journal of Phonetics, 5, 1977, 61-80

Delattre, P. Le jeu des transitions de formants et la perception des consonnes. Proceedings of the 4th International Congress of Phonetic Sciences, Helsinki. Den Haag: Mouton, 1962, 407-411

Dommelen, W. A. van. Assimilation of voice in German and Dutch. In M.van den Broecke & W.Zonneveld (Eds.), Sound Structures. Dordrecht: Foris Publ., 1983, 17-56

Donders, F. C. Over de tong-werktuigen van het stem- en spraakorgaan. Nederlandsch Archief voor Genees- en Natuurkunde, deel I, 1865, 451-472

Erickson, D, Baer, T. Harris, K. S. The role of the strap muscles in pitch lowering. In D.M.Bless & J.H.Abbs (Eds.), Vocal Fold Physiology: Contemporary research and clinical issues. San Diego: College-Hill Press, 1982, 279-285

Ewan, E. G. & Krones, R. Measuring larynx movement using the thyroumbrometer. Journal of Phonetics, 2, 1974, 327-335

Fischer-Jørgensen, E. Voicing, tenseness and aspiration in stop consonants, with special reference to French and Danish. Annual Report of the Institute of Phonetics, University of Copenhagen, 3, 1969, 63-114

Fischer-Jørgensen, E. Temporal relations in consonant-vowel syllables with stop consonants based on Danish material. In B.Lindblom and S.Ohman (Eds.), Frontiers of Speech Communication Research. New York: Academic Press, 1979, 51-68

Fischer-Jørgensen, E. Temporal relations in Danish tautosyllabic CV sequences with stop consonants. Annual Report of the Institute of Phonetics, University of Copenhagen, 14, 1980, 207-261

Fujimura, O. Physiological functions of the larynx in phonetic control. invited paper at the International Congress of Phonetic Sciences, Miami, 1977

Fujimura, O. & Miller, J. E. Mandible height and syllable-final tenseness. Phonetica, 36, 1979, 263-272

Gussenhoven, G. Voiced fricatives in Dutch: sources and present day usage. Proceedings of the Institute of Phonetics Nijmegen, 5, 1981 a, 84-95

Gussenhoven, G. Measuring the acceptability of voiced fricatives in Dutch. Proceedings of the Institute of Phonetics Nijmegen, 5, 1981 b, 96-129

Gussenhoven, C. & Bremmer Jr, R. H. Voiced fricatives in Dutch: Sources and present-day usage. North-Western European Language Evolution, 2, 1983, 55-71

Gussenhoven, C. & Broeders, A. The pronunciation of English: A course for Dutch learners. Groningen: Wolters-Noordhoff-Longman, 1976

Halle, M. & Stevens, K. N. On the feature "advanced tongue root". MIT Quarterly Progress Report, 94, 1969, 209-215

Halle, M. & Stevens, K.N. A note on laryngeal features. MIT Quarterly Progress Report, 101, 1971, 198-213

Hart, J. ´t & Cohen, A. Gating techniques as an aid in speech analysis. Language and Speech, 7, 1964, 22-39

Hirose, H. & Gay, T. The activity of the intrinsic laryngeal muscles in voicing control: An electromyographic study. Phonetica, 25, 1972, 140-164

Honda, K. Relationship between pitch control and vowel articulation. In D.M.Bless & J.H.Abbs (Eds.), Vocal Fold Physiology: Contemporary Research and Clinical Issues. San Diego: College-Hill Press Inc., 1983, 286-297

Hooff, C. M. C. van & Broecke, M. P. R. van den. Assimilation of voice in Dutch at three types of liguistic boundaries. Progress Report, Institute of Phonetics Utrecht, 8 no 1, 1983, 31-40

Hubers, G. A. C. & Kooij, J. G. Voice assimilation in Dutch. Acta Linguistica Hafniensa 14, 1973, 25-33

Jespersen, O. Lehrbuch der Phonetik (translation from Danish by H.Davidsen). Leipzig: B.G.Teubner Verlag, 1908

Jespersen, O. Elementarbuch der Phonetik, Dritter Hauptteil, Kombinationslehre. Leipzig: B. G. Teubner Verlag, 1912, 127-137

Kate, L. ten Onderzoek over onze Nederduitsche letterklanken (Re-editted by G.L.Meinsma: Als de haen over de heete kolen ?). Amsterdam: Institute of Phonetic Sciences, 1981

Keating, P. A. Comments on the jaw and syllable structure. Journal of Phonetics, 11, 1983, 401-406

Keating, P. A. Aerodynamic modeling at UCLA. UCLA Working Papers in Phonetics, 59, 1984, 18-28

Kohler, K. J. Experimental investigation into coarticulation and articulatory control in German. Arbeitsberichte Institut für Phonetik, Universität Kiel, 8, 1977

Kohler, K. J. Timing of articulatory control in the production of plosives. Phonetica, 38, 1980, 116-125

Lindqvist, J., Sawashima, M., Hirose, H. An investigation of the vertical movement of the larynx in a Swedish speaker. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo, 7, 1973, 27-34

Lisker, L. Rapid vs. Rabid: A catalogue of acoustic features that may cue the distinction. Haskins Laboratory Status Report on Speech Research, SR-54, 1978, 127-132

Lisker, L. & Abramson, A. S. A cross-language study of voicing in initial stops: acoustical measurements. Word, 20, 1964, 348-422

Lisker, L. & Abramson, A. S. Distinctive features and laryngeal control. Language, 47, 1971, 767-785

Loots, M. Syntax and assimilation of voice in Dutch. In M.van den Broecke & W.Zonneveld (Eds.), Sound Structures. Dordrecht: Foris Publ., 1983, 173-182

Meinsma, G. L. Assimilatie en assimilatieonderzoek. Taal en Tongval, 10, 1958, 107-114

Michels, L. C. Assimilatie, een kwestie van terminologie. Nieuwe Taalgids, 50, 1957, 210-212

Minifie, F. D., Abbs, J. H., Tarlow, A., Kwaterski, M. EMG activity within the pharynx during speech production. Journal of Speech and Hearing Research, 17, 1974, 497-504

Montanus, P. De Spreeckonst, 1635 (Re-editted by W.J.H.Caron). Groningen: J. B. Wolters, 1964

Muller, E. The effects of laryngeal articulation during stop consonant production on the time-course of supraglottal air pressure, in: D.M.Bless & J.H.Abbs (Eds.), Vocal Fold Physiology: Contemporary research and Clinical Issues. San Diego: College-Hill Press Inc, 1983, 166-178

Nooteboom, S. G. & Slis, I. H. The phonetic feature of vowel length in Dutch. Language and Speech, 15, 1972, 301-316

O'Connor, J. D. Phonetics. Hammondsworth, Middlesex: Penguin Books Ldt., 1973

Perkell, J. S. Physiology of speech production: Results and implications of a quantitative cineradiographic study. Cambridge, Mass.: The MIT-Press, 1969

Petursson, M. Aspiration et activité glottale; Examen expérimental à partir de consonnes islandaises. Phonetica, 33, 1976, 169-198

Pols, L. W. C. Variation and interaction in speech. To be published in J.Perkell et al. (Eds.), Proceedings of the Symposium on Invariance and Variability of Speech Processes. Hillsdale, N.J.: Lawrence Erlbaum Associates Inc., 1984 a

Pols, L. W. C. Phoneme identification in isolated stimuli and context. Proceedings of the Institute of Phonetic Sciences Amsterdam, 8, 1984 b, 33-40

Potter, R., Kopp, G., Green, H. Visible Speech. New York: Van Nostrand, 1947

Sawashima, M. Laryngeal research in experimental phonetics, In T.A.Sebeok (Ed.), Current Trends in Linguistics XII. The Hague: Mouton, 1974, 2303-2348

Sawashima, M. Laryngeal control for voicing distinctions: A review of recent works. Annual Bulletin of the Research Institute of Logopedics and Phoniatrics, University of Tokyo, 13, 1979, 23-26

Sawashima, M., Abramson, A. S., Cooper, F. S., Lisker, L. Observing laryngeal adjustments during running speech by use of a fiberoptics system. Phonetica, 22, 1970, 193-201

Schouten, J. F. On the perception of sound and speech. Invited paper IV-th International Congress on Acoustics Copenhagen, 1963, 195-207

Schouten, J. F. Subjective Stroboscopy and a model of visual movement detectors. Proceedings of the symposium on "Models of the Perception of Speech and Visual Form", Boston, 1967 a, 44-55

Schouten, J. F. Ten years perception research. IPO-Annual Progress Report, 2 (supplement), 1967 b, 7-16

Schouten, M. E. H. & Pols, L. C. W. Vowel segments in consonantal contexts: a spectral study of coarticulation. Journal of Phonetics, 7, 1979, 1-23

Slis, I. H. Articulatory measurements on voiced, voiceless and nasal consonants: A test of a model. Phonetica, 21, 1970, 193-210

Slis, I. H. Articulatory effort and its durational and electromyographic correlates. Phonetica, 23, 1971 a, 171-188

Slis, I. H. Rules for the synthesis of speech. IPO Annual Progress Report, 6, 1971 b, 28-34

Slis, I. H. On coarticulatory effort and syllable boundaries. IPO Annual Progress Report, 7, 1972, 40-44

Slis, I. H. Consequences of articulatory effort on articulatory timing. In G.Fant & M.A.A.Tatham (Eds.), Auditory Analysis and Perception of Speech. London: Academic Press, 1975, 397-412

Slis, I. H. Some remarks on speech synthesis by rule. Proceedings of the Institute of Phonetics Nijmegen, 2, 1978, 83-99

Slis, I. H. Synthesis by rule of three-consonant clusters. In Anniversaries in Phonetics: Studia Gratulatoria dedicated to Hendrik Mol. Amsterdam: Institute of Phonetic Sciences, 1979, 307-326

Slis, I. H. The effect of speaking rate on assimilation of voice. Proceedings of the Institute of Phonetics Nijmegen, 5, 1981 a, 160-176

Slis, I. H. The influence of the position of a stressed syllable on assimilation of voice. Proceedings of the Institute of Phonetics Nijmegen, 5, 1981 b, 177-187

Slis, I. H. Assimilatie van stem in het Nederlands. Glot, 5, 1982 a, 235-261

Slis, I. H. Voice quality and assimilation of voice. Proceedings of the Institute of Phonetics Nijmegen, 6, 1982 b, 53-60

Slis, I. H. Assimilation of voice in relation to voice quality. In M.van den Broecke & W.Zonneveld (Eds.), Sound Structures. Dordrecht: Foris Publ., 1983 a, 245-257

Slis, I. H. Assimilation of voice as a function of stress and word boundaries. Proceedings of the Institute of Phonetics Nijmegen, 7, 1983 b, 74-82

Slis, I. H. Assimilation of voice in Dutch. In M.P.R.van den Broecke & A.Cohen (Eds.) Prodeedings of the 10th International Congress of Phonetic Sciences, Utrecht. Dordrecht: Foris Punbl., 1984, 404-410

Slis, I. H. Assimilation of voice in Dutch. In R.Channon & L.Shocky (Eds.), In Honour of Ilse Lehiste / Ilse Lehiste Pühendusteots (accepted for publication). Dordrecht: Foris Publ., 1985

Slis, I. H. & Cohen, A. On the complex regulating the voiced-voiceless distinction I and II. Language and Speech, 12, 1969 a & b, 80-102 & 137-155

Slis, I. H & Muller, H. F. A computer programme for synthesis by rule. IPO Annual Progress Report, 6, 1971, 24-28

Slis, I. H., Nooteboom, S. G., Willems, L. F. Speech synthesis by rule: An overview of a system used in IPO. Hamburger Phonetische Beiträge, 22, 1977, 161-187

Sweet, H. The sounds of English. Oxford: Clarendon Press, 1908

Trommelen, M. & Zonneveld, W. Inleiding in de generatieve fonologie. Muiderberg: Coutinho, 1979

Vennard, W., Hirano, M., Ohala, J. Laryngeal synergy in singing, The NATS-Bulletin Vol. XXVII-1, 1970 a, 16-21

Vennard, W., Hirano, M., Ohala, J. Chest, Head and Falsetto, The NATS-Bulletin XXVII-3, 1970 b, 30-37

Viëtor, W. Elemente der Phonetik des Deutschen, Englischen und Französischen (4th ed.). Leipzig: O.R.Reisland, 1898

Weismer, G. & Fromm, D. Acoustic analysis of geriatric utterances: segmental and nonsegmental characteristics that relate to laryngeal function. In D.M.Bless & J.H.Abbs (Eds.), Vocal Fold Physiology: Comtemporary Research and Clinical Issues. San Diego: College-Hill Press, 1983, 317-332

Willems, L. F. IPOVOX-II: A speech synthesizer. IPO Annual Progress Report, 1, 1966, 120-125

Wingate, A. H. A phonetic answer to a phonological question. UCLA Working Papers in Phonetics, 54, 1982, 1-27

Zwaardemaker Cz, H. & Eijkman, L. P. H. Leerboek der Phonetiek. Haarlem: De Erven Bohn, 1928

CURRICULUM VITAE

Iman Slis werd op 21 april 1932 in Bandoeng op het eiland Java geboren. Na het behalen van het diploma HBS-b aan het Hervormd Lyceum te Amsterdam in 1950, studeerde hij pharmacie aan de Universiteit van Amsterdam. In 1955 brak hij deze studie af om na het vervullen van de militaire dienstplicht in 1957 in dienst te treden van de N.V. Philips te Eindhoven. Tot 1962 werkte hij daar op het Fysisch-Technisch Röntgen Laboratorium aan de bepaling van röntgen-beeldkwaliteiten. In deze periode volgde hij de bedrijfscursussen voor fysisch assistent en natuurkunde. Van 1962 tot 1976 was hij voor de zelfde firma verbonden aan het Instituut voor Perceptie Onderzoek (IPO) te Eindhoven waar hij werkte aan verschillende onderwerpen op het terrein van het fonetisch onderzoek. Het centrale thema daarbij was spraaksynthese. In 1962 behaalde hij een propedeutisch getuigschrift in de technische natuurkunde aan de Technische Hogeschool te Eindhoven. In 1973 behaalde hij de diplomas logopedie en akoepedie, waarna hij als part-time docent aan de logopedieopleiding verbonden is geweest; eerst in de fonetiek aan de opleiding te Eindhoven, en later in de fonetiek en de audiologie in Nijmegen. In 1975 is hij een half jaar bij het Philips Project Centre in Geldrop gedetacheerd geweest voor onderzoek aan een spraaktechnologisch onderwerp. Sinds 1976 is hij als wetenschappelijk medewerker verbonden aan het Instituut voor Fonetiek van de Katholieke Universiteit te Nijmegen, waar hij in 1983 een getuigschrift "vrij doctoraal letteren" verkreeg.

STELLINGEN BEHOREND BIJ HET PROEFSCHRIFT VAN I. H. SLIS:
THE VOICED-VOICELESS DISTINCTION AND ASSIMILATION OF VOICE IN DUTCH

1. Stemstoornissen kunnen het best bestudeerd worden in het spraaksignaal dat grenst aan de overgangen van stemhebbende naar stemloze en van stemloze naar stemhebbende stukken.

2. De betrekkelijk grote verschillen in geobserveerde assimilaties van stem, zoals die in de fonetische literatuur van voor 1960 te vinden zijn, worden in belangrijke mate veroorzaakt door een persoonlijke bias van de transcribenten.

3. De perceptieve bijdrage van het stemsignaal in de gesloten periode (occlusiefase) van een stemhebbende medeklinker is relatief klein ten gevolge van de voorwaartse en achterwaartse maskering van de omgevende klinkers.

4. De duur van medeklinkers levert een belangrijke bijdrage aan de waarneming van syntactische grenzen.

5. Het gegeven dat de duur van spraakklanken (foneemrealisaties) in gefluisterde spraak langer is dan in normale spraak, zou geinterpreteerd kunnen worden als een gevolg van een grotere articulatorische inspanning bij fluisteren.

6. Dat eventuele nasaliteit in gesloten klinkers slecht waarneembaar is wordt veroorzaakt door het feit dat de voor nasaliteit kenmerkende formant onder de 500 Hz ongeveer samenvalt met de eerste formant van die klinkers.

7. Het bepalen van de eerste formant van met stem gesproken klinkers wordt (vooral bij vrouwenstemmen) bemoeilijkt door relatief grote frequentieverschillen van de harmonischen in deze eerste formant; dit probleem zou ondervangen kunnen worden door deze formant bij gefluisterde spraak te meten.

8. Modellen, die gebaseerd zijn op woord-aangroei-proeven, en waarbij aangenomen wordt dat de spraakherkenning sequentieel (van links naar rechts) plaatsvindt (bottum-up modellen), zijn in strijd met het experimentele gegeven dat luisteraars in een woord onhoorbaar gemaakte klanken menen waar te nemen; klaarblijkelijk vult een luisteraar ontbrekende informatie aan uit eigen kennis van de taal (top-down modellen) zonder dat hij zich ervan bewust is dat het herkenningsproces op het moment van de onhoorbaar gemaakte klank verstoord wordt.

9. Docenten in het uitspraakonderwijs zouden een gedegen kennis van de fonetiek moeten hebben.

10. De nogal eens gehoorde bewering dat de tijd van een arts geld kost mag nooit als rechtvaardiging gebruikt worden voor lange wachttijden van patienten in volle wachtkamers, wat immers de gemeenschap nog veel meer geld kost.

11. Het kweken van groente en fruit in eigen tuin kweekt bovendien een milder standpunt ten aanzien van regenachtig weer.